Retroviral-RNA Structure and Function:

Investigating the role of aminoacyl-tRNA synthetases and retroviral-RNA structural elements in the initiation of reverse transcription

Honors Research Thesis

Presented in partial fulfillment of the requirements for graduation with honors research distinction in Biochemistry in the undergraduate colleges of The Ohio State University

By

Joshua E. Hatterschide

The Ohio State University
April 2016

Project Advisor: Dr. Karin Musier-Forsyth, Department of Chemistry and Biochemistry

Table of Contents

Abbreviations 2
Introduction 3
Chapter 1: RNA Probing Data Analysis Method 5
Introduction 5
Materials and Methods 8
RNA preparation 8
5′ fluorescently labeling reverse transcription primer 9
5′ ^{32}P-labeling reverse transcription primer 9
Sequencing 9
SHAPE experiments 10
Ribonuclease optimization experiments analyzed via PAGE 10
Ribonuclease digestion experiments analyzed via CE 11
Results and Discussion 11
Section I: RNA probing with RNases and SHAPE 11
Section II: RNA probing data analysis improvements 14
Conclusions and Future Directions 22
Acknowledgements 23
Chapter 2: Human Immunodeficiency Virus Type I 24
Introduction 24
Materials and Methods 26
RNA preparation 26
Protein preparation 27
Small-angle X-ray scattering 28
Ribonuclease/SHAPE protection assays 28
Sequencing 28
Results and Discussion 29
Section I: LysRS interaction with HIV-1 5′UTR(356) probed by RNases and SHAPE 29
Section II: SAXS study of HIV-1 5′UTR ΔTAR/polyA 32
Conclusions and Future Directions 36
Acknowledgements 37
Chapter 3: Human T-cell Leukemia Virus Type I 38
Introduction 38
Materials and Methods 40
RNA Preparation 40
Protein Preparation 41
Electrophoretic mobility shift assay 41
Fluorescent RNA labeling 42
Fluorescent Protein Labeling 42
Fluorescence anisotropy direct binding assay 42
Results and Discussion 42
Conclusions and Future Directions 47
Acknowledgements 48
Conclusion 49
References 51

Abbreviations

AIDS	Acquired immunodeficiency syndrome
ATL	Adult T-cell leukemia
CE	Capillary electrophoresis
DIS	Dimerization initiation site
EMSA	Electrophoretic mobility shift assays
EPRS	Glutamyl-prolyl-tRNA synthetase
FA	Fluorescence anisotropy
GAIT	Gamma-activated inhibitor of translation
HAM/TSP	HTLV-1 associated myelopathy/tropical spastic paraparesis
HIV	Human immunodeficiency virus
HTLV	Human T-cell lymphotropic virus
LysRS	Lysyl-tRNA synthetase
MW	Molecular weight
NMIA	N-methylisatoic anhydride
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PBS	Primer binding site
PDDF	Pair-distance distribution functions
ProRS	Prolyl-tRNA synthetase
RNase	Ribonuclease
RSV	Rous sarcoma virus
RT	Reverse transcription
SAXS	Small-angle X-ray scattering
SEC	Size exclusion chromatography
SHAPE	Selective 2'-hydroxyl acylation analyzed by primer extension
SIV	Simian immunodeficiency virus
TAR	Trans activating response
TLE	tRNA-like element
WHO	World Health Organization
WT	Wild type

Introduction

Retroviruses are a diverse family of viruses with members that affect nearly all vertebrates, and many of them carry oncogenes that cause cancerous growth in their host organisms. Since the discovery of the first human retrovirus in a patient with a cutaneous T-cell lymphoma in 1979 (Gallo, 2005), retroviruses have been found to be the cause of devastating diseases including acquired immunodeficiency syndrome (AIDS) and adult T-cell leukemia (ATL) that affect millions of people worldwide. Many of these diseases still have no treatment, and none have been cured.

The *retroviridae* family is divided into seven different genera, three of which contain viruses that infect humans: lentiviruses, deltaretroviruses, and spumaviruses. However, the only spumavirus to be isolated from humans, prototype foamy virus, is non-transmissible and nonpathogenic in humans (reviewed by Meiering et al., 2001). Although most retroviruses from the other two genera are transmissible from human to human, the majority of them are not known to be pathogenic. In fact, the only two retroviruses that have been shown to cause human disease are human immunodeficiency virus (HIV), and human T-cell lymphotropic virus type 1 (HTLV-1)—the viruses that cause the diseases mentioned above and are consequently the topic of this thesis.

Retroviridae differs from many other viral families due to two major lifecycle events: reverse transcription (RT) and integration (Figure 1). During RT, the viral RNA genome is copied from ssRNA to dsDNA after which it can be imported into the nucleus to undergo integration into the host cell's genome. For RT to initiate, a primer must specifically anneal to the primer binding site (PBS) in the 5' untranslated region (5'UTR) of the retroviral genome. All retroviruses and long terminal repeat retrotransposons use host cell tRNAs as the primer for reverse transcription (Marquet et al., 1995; Mak et al., 1997), including the two viruses of interest: HIV-1 uses $tRNA^{Lys3}$, and HTLV-1 uses $tRNA^{Pro}$. All three human $tRNA^{Lys}$ isoacceptors along with human lysyl-tRNA synthetase (LysRS) are selectively packaged into HIV-1 particles (Jiang et al., 1993). Our lab has demonstrated that human LysRS facilitates priming in HIV-1 by binding a

tRNA-like-element (TLE) near the PBS, localizing the $tRNA^{Lys3}$ primer (Jones et al., 2013; Jones & Cantara et al., 2014). As the primer for HTLV-1 RT is $tRNA^{Pro}$, we hypothesized that this virus could employ a similar mechanism involving human glutamyl-prolyl-tRNA synthetase (EPRS) binding to a TLE.

The lifecycle of a retrovirus is highly complex, and the 5'UTR is at the center of more than just primer localization (Figure 1). Therefore, although we are particularly interested in the details of how aminoacyl-tRNA synthetases are potentially involved in packaging the primer that is required for RT, we are also interested in the structure and function of this large RNA as a whole. In this thesis, I will discuss improvements made to the analysis of RNA probing data, experiments performed to study the structure and function of the HIV-1 5'UTR, and investigations of the putative interaction between elements within the HTLV-1 5'UTR and EPRS. Using a wide variety of biochemical techniques, we hope to elucidate information about these viral processes that could eventually aid in the production of therapeutic agents.

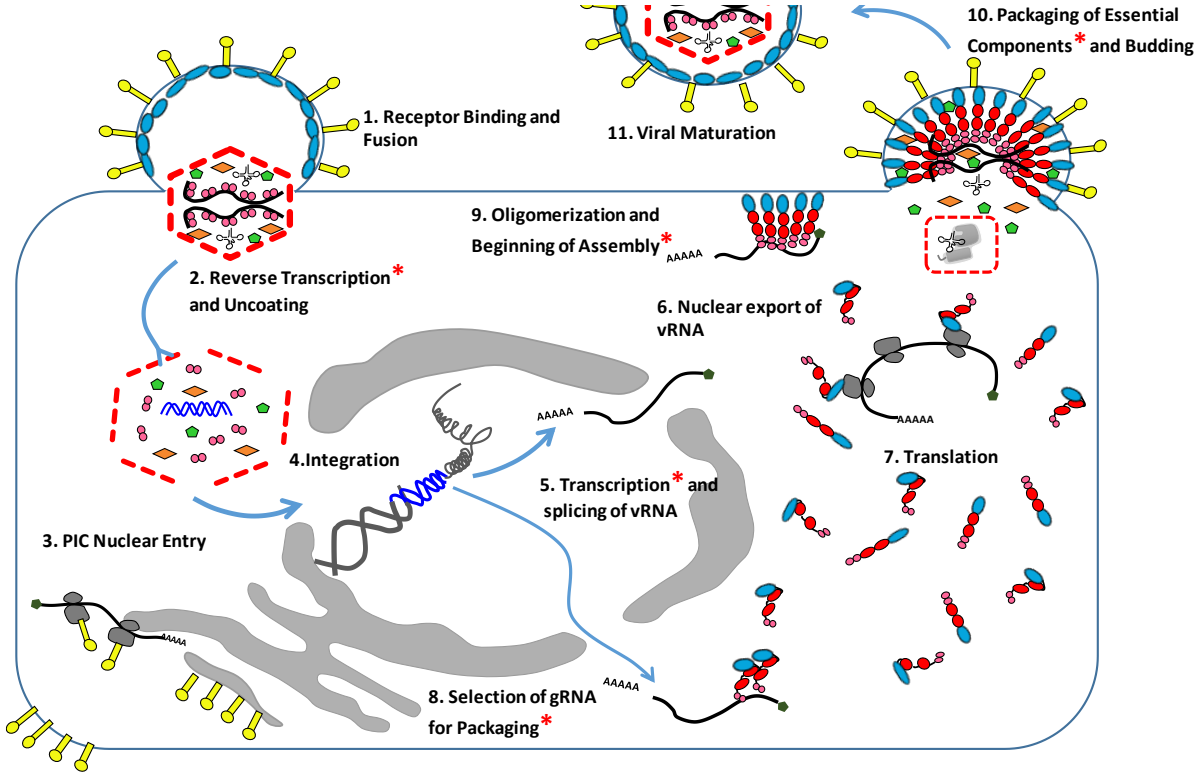


Figure 1: The retroviral life cycle (modified from Dr. Tiffiny Rye-McCurdy). RT and integration can be seen in steps 2 and 4 of the viral lifecycle. The packaging of the tRNA primer and aminoacyl tRNA synthetase are shown in the red dashed box. Steps of the viral lifecycle that directly involve the 5'UTR in the viral RNA or proviral DNA are marked by a red "*".

Chapter 1: RNA Probing Data Analysis Method

Introduction

In addition to serving as a key player in translation of the genetic code, RNA is capable of having many regulatory and enzymatic functions as well. As with proteins, these functions are directly related to the three dimensional (3D) shape, or tertiary structure, of the RNA molecule. However, crucial to understanding the 3D structure and function of an RNA is knowledge of the base-pairing interactions, or secondary structure, of the molecule. In conjunction with informing the tertiary structure, the secondary structure can aid in the search for familiar RNA functional motifs and facilitate the design of RNA constructs that are likely to maintain their native fold. However, the current RNA secondary structure-prediction algorithms are not usually sufficient for determining the fold of large RNA molecules, such as retroviral genomes, because they are unable to distinguish between structures with similar free energies (Rice et al., 2014). Furthermore, traditional structural techniques such as X-ray crystallography and NMR often fall short as well, due to the flexibility of many RNA molecules in solution and their slow rate of rotation in solution (Cantara & Olson et al., 2014). Therefore, using chemical and enzymatic agents to probe the structures of these molecules is essential to obtaining accurate structures of most RNAs (Figure 2A) (Weeks, 2010). Additionally, these techniques can be useful in studying RNA-protein complexes, as they can be used to interrogate the location of protein binding to the RNA molecule, as well as any secondary structural changes this binding confers to the RNA (Figure 2B) (Xu et al., 2009).

There are a large number of RNA probing methods, but two will be the focus of the current chapter: ribonuclease (RNase) digestion and selective 2’ hydroxyl acylation analyzed by primer extension (SHAPE)—each with its own set of benefits and drawbacks. RNase digestion involves reacting RNAs with enzymes that cleave the RNA backbone with specific nucleotide and base-pairing preferences. From analysis of the digestion pattern, whether a nucleotide is in a single- or double-stranded region can be determined. Although primer extension is not required for the analysis of RNase digestion experiments, it can be a useful tactic; since the resulting cDNA is much more stable than the initial RNA, it improves the

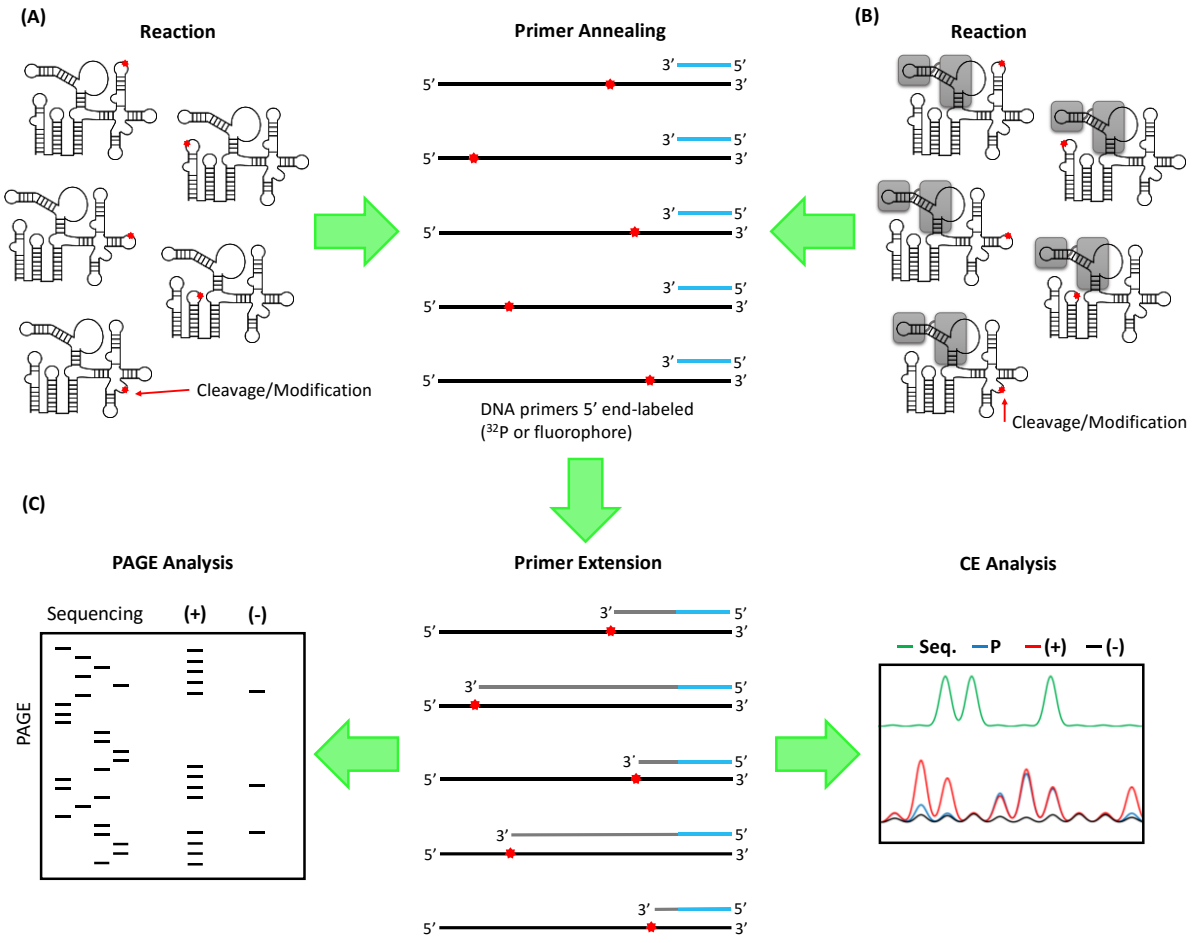


Figure 2: Experimental schematic for RNA structure probing (A) and RNA-Protein footprinting (B). For either technique RNAs are subjected to reactions with enzymes or chemical probes that result in cleavages to the RNA backbone or covalent modifications. The resulting RNA from the reactions is then denatured, and 5'end-labeled primers are annealed for detection after reverse transcription. During primer extension, reverse transcriptase will not be able to extend past cleavages or modifications resulting in cDNA of varying lengths. (C) Primer extension products can be analyzed by either PAGE (left) or CE (right). "Seq." stands for sequencing, "P" stands for protection, "(+)" represents the reaction, and "(-)" represents the background.

analysis of longer RNAs, and it allows these reactions to be analyzed in a high-throughput manner. RNase digestion can be advantageous for multiple reasons: (1) it is a robust enzymatic reaction, (2) it is one of the simplest probing reactions that is capable of resolution in double-stranded regions of RNA, and (3) RNases are sterically large agents that can be useful for RNA-protein footprinting. However, due to enzyme specificities, multiple enzymes are required to obtain resolution throughout the RNA, making this potentially more labor intensive than other methods. SHAPE on the other hand, uses reactive anhydrides

to selectively acylate the 2' hydroxyl group in the ribose sugar of the RNA backbone. The kinetics of this reaction are dependent on the flexibility of the RNA at each nucleotide (Merino et al., 2005). Thus, as single-stranded RNA is much more flexible, these nucleotides will also be the most reactive to SHAPE reagents. During primer extension, reverse transcriptase cannot move past the SHAPE-modified sugar in the RNA backbone, allowing for the quantification of the reactivity at each nucleotide (Merino et al., 2005). The major advantage of SHAPE is that it is capable of detecting nearly every single-stranded and/or flexible nucleotide, making it extremely useful for structure probing. Conversely, low reactivity at base-paired nucleotides makes SHAPE less useful for RNA-protein footprinting. There are also many other RNA probing techniques involving base-specific chemical reactions, or hydroxyl radical footprinting (reviewed in Weeks, 2010); however, these will not be discussed here.

After the reactions and primer extensions are complete there are also three different methods of analyzing the results: (1) urea polyacrylamide gel electrophoresis (PAGE), (2) capillary electrophoresis (CE), and (3) next generation sequencing. Both CE and next generation sequencing are considered high-throughput, meaning they allow for many more experiments to be more accurately and easily quantified than PAGE. Next generation sequencing is much less cost effective than CE, so our efforts were focused on CE. CE data can be difficult to interpret when trying to optimize techniques. Therefore, the first section of this chapter will demonstrate how PAGE analysis was used to optimize RNA probing and sequencing experiments, leading to the qualitative analysis of CE data.

The data CE produces must be heavily processed before they can be accurately analyzed, and this processing of high-throughput RNA probing is what we are interested in improving. There are currently a few programs that exist that facilitate CE data analysis: SHAPEfinder (Vasa et al., 2008), QuShape (Karabiber et al., 2013), and FAST (Pang et al., 2011). However, each of these has disadvantages. None of these programs are transparent, meaning they don't allow you to see and change the calculations that the program is performing. Second, all three of these are relatively slow, frequently make errors, and do

not possess easy controls for correcting these errors. The second section of this chapter will present our modifications to the methods that are used in the currently accepted programs; it will also introduce the transparent tool that we have developed in order to facilitate our calculations. This method is pending validation by reproducing the previously determined secondary structure of the HIV-1 5'UTR.

Materials and Methods

RNA preparation

The 105 nt HIV-1 PBS/TLE RNA (Jones et al., 2013) was used for all RNA probing optimization experiments analyzed by PAGE. This RNA was previously cloned out of the HIV-1 5'UTR, and into a pUC19 vector behind a T7 promoter (Jones et al, 2013). The transcription template was obtained by digestion of the plasmid with FokI restriction endonuclease. FokI was chosen because its cut site is distant from its recognition site, so it does not restrict the sequences that can be chosen for the 3' end of the RNA construct (Hiroyuki et al., 1981).

Three RNA constructs were analyzed in SHAPE/CE experiments—HIV-1 5'UTR(356) WT, HIV-1 5'UTR(356) ΔDIS, and HIV-1 5'UTR(356) ΔDIS A34U. The HIV-1 5'UTR(356) wild type (WT) construct contains the first 356 nucleotides of the NL4-3 isolate. The ΔDIS mutation was previously created from the pHIV-1 5'UTR(356) WT plasmid by mutating the dimerization initiation site (DIS) loop to a stable GAGA tetraloop (E. Olson, unpublished data). The A34U mutation was also previously created from the pHIV 5'UTR(356) ΔDIS plasmid by mutating A34 to T in the *trans* activating response element (TAR). Both the A34T and ΔDIS mutations prevent genomic dimerization and facilitate homogeneous RNA preparation (Helga-Maria et al., 1999; Skripkin et al., 1994). These plasmids were then digested with FokI to create the transcription template.

After obtaining transcription templates, RNAs were prepared via *in vitro* transcription with T7 RNA polymerase (Milligan et al. 1987), and analyzed for homogeneity using 8M urea (denaturing) PAGE. Desired bands were excised, crushed, and soaked in RNA elution buffer (0.5 mM NH_4OAc, 1 mM EDTA)

overnight in a 37°C shaker. The supernatant was then butanol extracted and ethanol precipitated to yield a pellet of purified RNA. RNAs were folded in 50 mM HEPES (pH 7.4) buffer by heating at 80°C for 2 minutes, cooling to 60°C for 2 minutes, adding 1 M $MgCl_2$ to a final concentration of 10 mM, and (if longer than ~150 nt) incubating at 37°C. The three HIV-1 5'UTR(356) RNA constructs were optimized for folding homogeneity by varying the 37°C incubation time.

5′ fluorescently labeling reverse transcription primer

To analyze RNA probing via CE, primers need to be labeled with particular fluorophores that can be detected. Primers for reverse transcription and cycle sequencing 5'-end labeled with NED® dye were ordered from IDT.

5′ ^{32}P-labeling reverse transcription primer

Reverse transcription and sequencing primers for the PBS/TLE were 5'-end labeled with ^{32}P using T4 polynucleotide kinase (PNK) and γ-^{32}P ATP. Reactions were incubated for 1 h at 37°C and then gel purified by denaturing PAGE. Bands were excised, crushed and soaked in elution buffer at room temperature overnight. The supernatant was then ethanol precipitated to yield pure ^{32}P-labeled primers. These were then quantified using a liquid scintillation counter.

Sequencing

Sequencing was performed using the Thermo Sequenase Cycle Sequencing kit (Affymetrix), a Sanger style sequencing kit utilizing dideoxynucleotides (ddNTPs). NED- or ^{32}P-labeled primers were designed to anneal to the part of the plasmid corresponding to the 3' end of the RNA. Primer extension was performed in the presence of one of the four ddNTPs, resulting in the termination of primer extension with every ddNTP addition. These sequencing reactions were then analyzed by PAGE and CE to provide a ladder that can be aligned with the probing data.

SHAPE experiments

Two different SHAPE reagents were used in this study: N-methylisatoic anhydride (NMIA) (Sigma-Aldrich) and 1-methyl-6-nitroisatoic anhydride (1M6) (Sigma-Aldrich). NMIA reacts slowly, and is useful for characterizing nucleotides that exhibit slow dynamics (Rice et al., 2014). In contrast, 1M6 reacts more rapidly and also deactivates with water much quicker making it more useful for examining nucleotides with faster dynamics (Rice et al., 2014). Prior to experimental data collection, RNA reactions were time optimized with SHAPE reagents to ensure single-hit kinetics were obtained. SHAPE experiments were performed with a plus reaction (with SHAPE reagent) and a minus reaction (without SHAPE reagent) to control for background effects such as spontaneous reverse transcription termination. Reactions were initiated with the addition of 1 µL of 80 mM SHAPE reagent to the plus reaction and 1 µL DMSO to the minus reaction, each tube containing approximately 8 pmols of RNA in 9 µL buffer (final reaction volume of 10 µL). Reactions were then incubated at room temperature for the optimized time duration (22 min. for NMIA, and 3 min. for 1M6). Reactions were then quenched by ethanol precipitation. RNA pellets were resuspended and reverse transcribed with NED-labeled primers using Superscript III reverse transcriptase by following the manufacturer's suggested protocol (Invitrogen).

Ribonuclease optimization experiments analyzed via PAGE

Optimization of RNase T1 (which cleaves at single stranded guanosine residues) digestion conditions was performed on the HIV-1 PBS/TLE RNA using three different reactions: structured probing, unstructured probing, and alkaline hydrolysis. Structured probing was carried out by the addition of RNase T1 to folded RNA followed by incubation at room temperature for 15 minutes. The reactions were then quenched with precipitation/inactivation buffer (Ambion), and RNA pellets were reverse transcribed with ^{32}P-labeled primers using Superscript III (Invitrogen). The goal of this reaction is to determine which G residues are single stranded. For unstructured probing, RNAs were heated to 50°C for 5 min to unfold them prior to addition of RNase T1. The remaining steps were carried out identically to structured probing.

The goal of this reaction is to provide a ladder of all of the G residues in the RNA. Alkaline hydrolysis was performed by incubation of RNA at 95°C in alkaline hydrolysis buffer (Ambion). Reactions were then quenched and carried out identically to the other two reactions. The goal of this reaction is to provide a ladder of every nucleotide of the RNA. Structured and unstructured reactions were optimized by the addition of varying amounts of RNase T1 (0.001 to 0.1 units/μL) and alkaline hydrolysis was optimized for incubation time (5 to 15 min). These experiments were analyzed by denaturing PAGE and phosphorimaging of the gel.

Ribonuclease digestion experiments analyzed via CE

Two RNases with different strand specificities and base preferences were used the RNA: RNase T1 and RNase V1. These RNases prefer single-stranded guanosines and double-stranded nucleotides respectively. All experimental data were collected with a plus and a minus reaction. To initiate the reaction, RNases were added to folded RNAs in the plus tube and buffer was added to the minus. They were then incubated at room temperature for 20 min and quenched by precipitating the RNA with premixed precipitation/inactivation buffer (Ambion). RNA pellets were resuspended and reverse transcribed with NED-labeled primers using Superscript III.

Results and Discussion

Section I: RNA probing with RNases and SHAPE

RNA probing experiments are capable of providing secondary structure information at single-nucleotide resolution. However, these experiments can be complex, requiring optimization of several steps before the results can be accurately analyzed: such as folding conditions, reaction times, primer extension reactions, and sequencing. The probing reaction needs to be optimized to obtain single-hit kinetics. Failure to do so will result in a bias toward reactivity closer to the 3' end of the RNA as reverse transcriptase will halt at the first cleavage/modification that it encounters. Additionally, the quantity of

primer extension product and sequencing product that are optimal for analysis must be determined. The remainder of this section will focus on description of how RNase digestion, primer extension, and sequencing experiments were optimized using PAGE, and how a new tool was developed to facilitate quicker and more user-friendly analysis of the CE data.

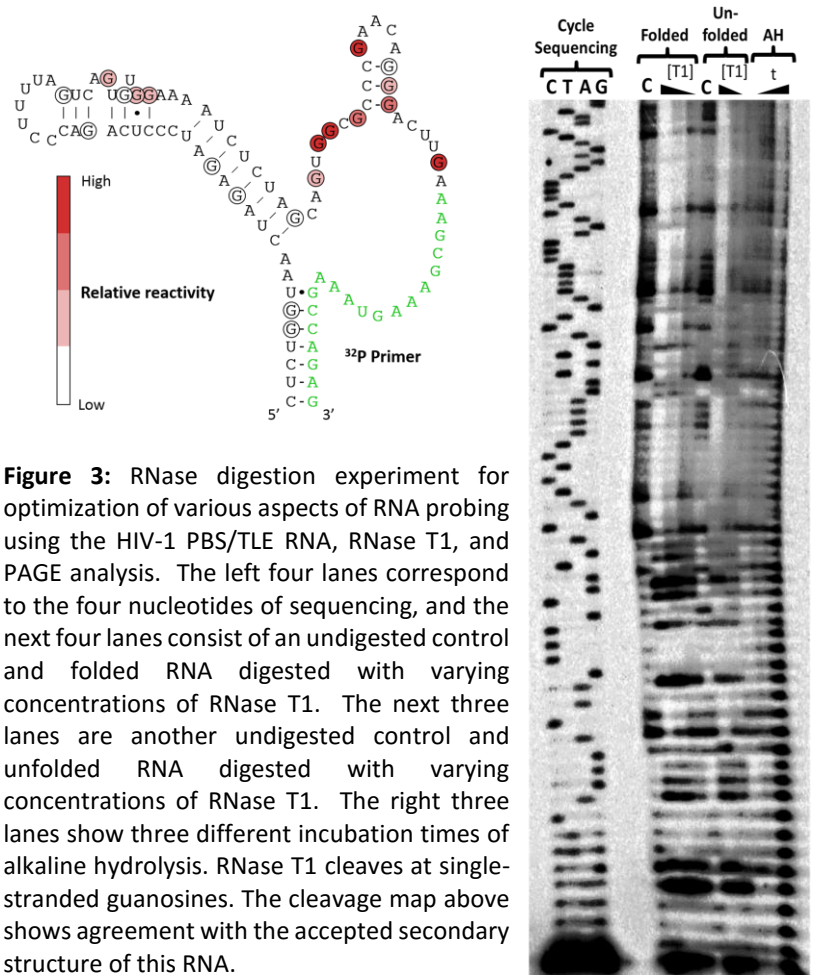


Figure 3: RNase digestion experiment for optimization of various aspects of RNA probing using the HIV-1 PBS/TLE RNA, RNase T1, and PAGE analysis. The left four lanes correspond to the four nucleotides of sequencing, and the next four lanes consist of an undigested control and folded RNA digested with varying concentrations of RNase T1. The next three lanes are another undigested control and unfolded RNA digested with varying concentrations of RNase T1. The right three lanes show three different incubation times of alkaline hydrolysis. RNase T1 cleaves at single-stranded guanosines. The cleavage map above shows agreement with the accepted secondary structure of this RNA.

RNase optimization experiments (see Methods and Materials for detailed protocol) were chosen for initial experiments in our lab because they involve simple enzymatic reactions that can be analyzed both directly via denaturing PAGE and using primer extension analyzed by CE. After initial control experiments to ensure proper function of the RNase and polymerase enzymes, experiments were undertaken to optimize the experimental concentration of RNase T1 (Figure 3). These experiments also served to validate our reaction, primer extension, and sequencing conditions by demonstrating the correct digestion pattern of the HIV-1 PBS/TLE RNA. Despite this success, further PAGE-analyzed experiments were not continued due to their many limitations. First, the window of analyzable data on a sequencing PAGE gel is very small, limiting experiments to small RNAs unless many reverse transcription primers are used. Second, sequencing PAGE gels are very delicate and are frequently damaged while

fixing or drying, resulting in loss of data. Finally, these experiments require the use of radioactivity, necessitating extra safety precautions during experimentation. For these reasons, we proceeded to analyze experiments via CE.

The RNase digestion experimental method was simplified significantly when translated to CE analysis. Experiments involving RNase reactions with the unfolded RNA were eliminated due to the inclusion of the much more reliable sequencing lane. During reverse transcription, a small fraction of the polymerase population will spontaneously terminate at each nucleotide. While the resulting fragments from these terminations are difficult to detect via PAGE, they can be easily resolved using CE. As a result, the single-nucleotide ladder provided by alkaline hydrolysis was redundant and removed from the protocol as well. Therefore, RNase experiments analyzed by CE involved three types of samples: RNase digestions, background (no RNase), and sequencing.

Additionally, it is possible to resolve much larger RNAs using CE than PAGE, so full-length 5'UTR constructs were used in these experiments. Homogeneity of these constructs was tested using native PAGE (see materials and methods for detailed protocol). It was found that the HIV-1 5'UTR(356) ΔDIS A34U construct was the most homogenous (Figure 4). As a result, this RNA was selected for use in RNA probing experiments analyzed by CE.

SHAPE probing was also optimized with the HIV-1 5'UTR(356) ΔDIS A34U. Single-hit kinetics were optimized by testing multiple incubation times for two reagents, 1M6 and NMIA. Single-hit kinetics are necessary because reverse transcriptase will always terminate at the first modification that it encounters. If there is more than one modification on a

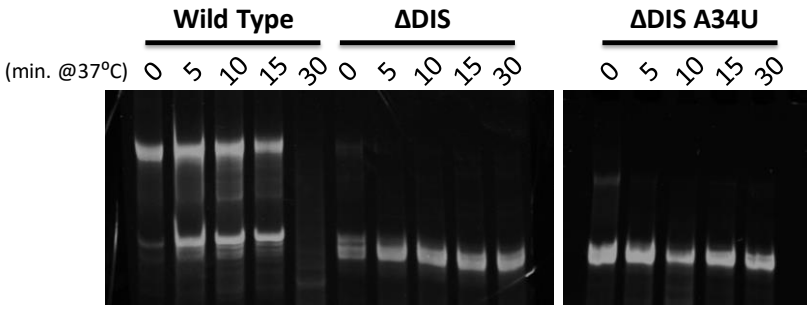


Figure 4: Three different HIV-1 5'UTR constructs were optimized for folding by varying the incubation time at 37°C and analyzed using native PAGE. The 5'UTR ΔDIS A34U incubated at 37°C for 30 min. was the most homogenous.

given RNA, only the first one will be detected resulting in a bias in reactivity toward the 3' end. Therefore, the criteria for optimal incubation was the time that showed the lowest 3'-end bias. However, all of the incubation times tested gave relatively even distributions of highly reactive nucleotides and little to no bias toward the 3' end was observed (Figure 5). This was likely the result of water inactivation of the anhydride. As a result, the shortest incubation times were chosen arbitrarily. In subsequent experiments, both structure probing and protection, a high rate of spontaneous termination of reverse transcription was observed, resulting in low processivity of the polymerase and many negative reactivity values. This may be due to the highly-structured nature of the HIV-1 5'UTR. Experiments are in progress to test different thermo-denaturing cycles to improve reverse transcriptase processivity.

	1M6 Optimization		
	3 min	**10 min**	**15 min**
# Highly Reactive Peaks	31.0	26.0	24.0
% 3' Half	54.8%	50.0%	45.8%
% 5' Half	45.2%	50.0%	54.2%
Difference (3'-5')	9.7%	0.0%	-8.3%

	NMIA Optimization		
	22 min	**30 min**	**45 min**
# Highly Reactive Peaks	23.0	31.0	34.0
% 3' Half	52.2%	48.4%	44.1%
% 5' Half	47.8%	51.6%	55.9%
Difference (3'-5')	4.3%	-3.2%	-11.8%

Figure 5: HIV-1 5'UTR ΔDIS A34U RNA was optimized for single-hit kinetics with 1M6 and NMIA. All incubation times for both reagents resulted in even distribution of highly reactive peaks, and no significant preference for the 3' end of the RNA molecule.

Section II: RNA probing data analysis improvements

There are three programs that are predominantly used for analysis of RNA probing data—two programs produced by the Weeks lab: SHAPEfinder (Vasa et al., 2008) and QuShape (Karabiber et al., 2013), and a program produced by the Glenn lab called Fast (Pang et al., 2011). There are also other, less commonly-used, programs that will not be discussed in detail. Therefore, I will primarily focus on how our method addresses issues with FAST and QuShape. Our analysis method can be divided into four major steps: signal alignment, preprocessing, sequence alignment, and reactivity calculation. Each of these steps consists of multiple calculations (Figure 6) and how they improve upon other programs will be discussed.

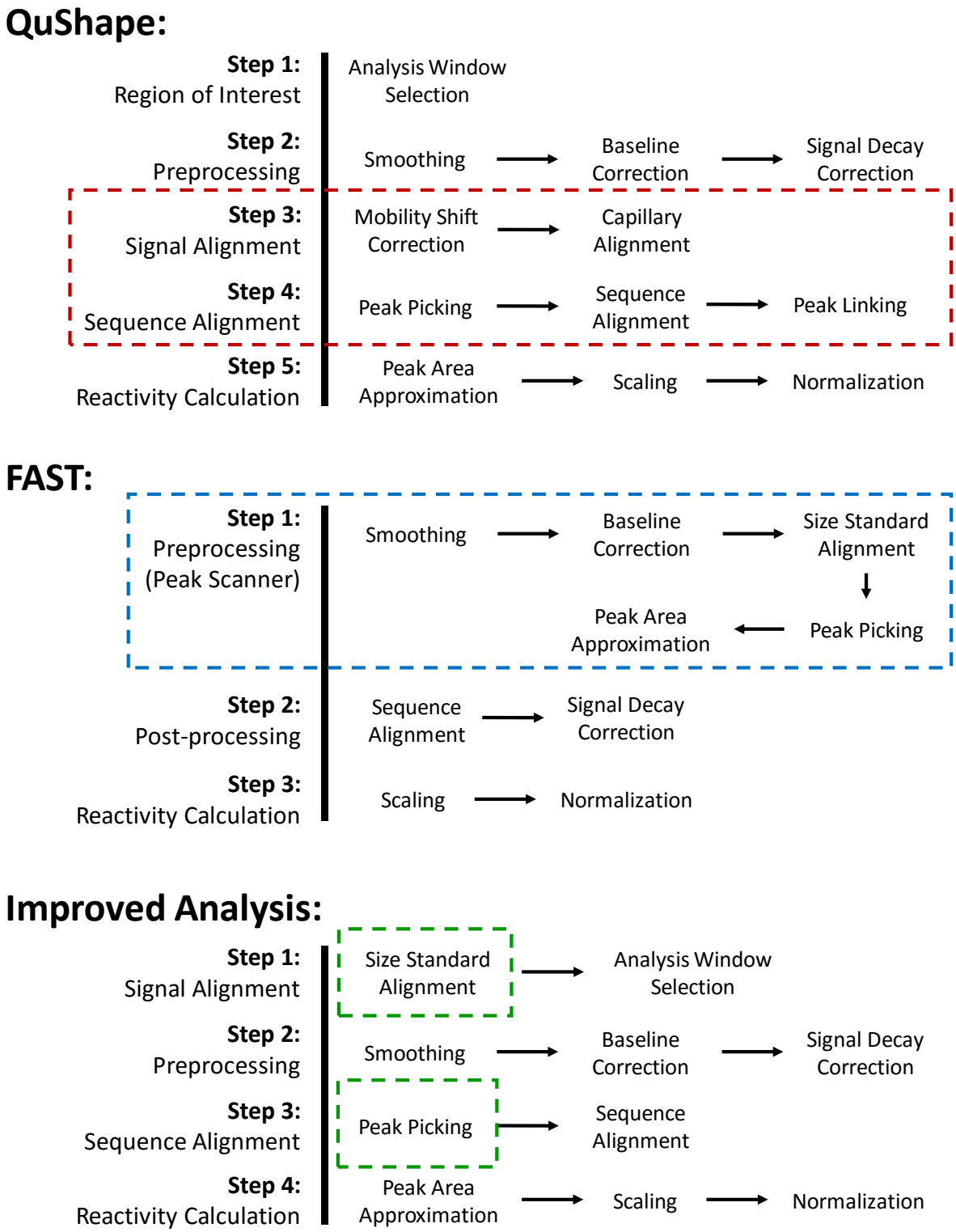


Figure 6: The processing flow charts for QuShape, FAST, and our improved analysis method are depicted above. The red-dashed box indicates the processing steps that frequently require cumbersome manual override in QuShape. The blue-dashed box indicates the processing steps of FAST that have little transparency due to their dependence on Peak Scanner. The green dashed boxes indicate the primary calculations in our improved analysis method that address these issues.

There are three aspects of QuShape that we address with our method: signal alignment, sequence alignment, and user interface. There are two calculations involved in the signal alignment step: mobility shift and capillary alignment. QuShape requires experiments to be set up in a two capillary, two dye scheme: reactions and sequencing run in the same capillary with different dyes, and the (+) and (-) reactions in different capillaries (Karabiber et al., 2013). The idea behind this is that both reaction and sequencing can be run in the same capillary and therefore migrate under the same conditions. However, this is not exactly the case because the dye has an effect on the migration rate of the DNA fragment; thus, this setup requires a mobility-shift correction to account for different dye mobilities. This does not significantly affect accuracy or analysis time, but is an aspect that can be eliminated by experimental set up. Additionally, as QuShape does not utilize size standards, it uses a mathematically complex algorithm to align the electropherograms between capillaries using a similarity matrix (Karabiber et al., 2013). This algorithm often misaligns CE traces requiring a time-consuming manual override. Incorporating size standards into the experimental set up (discussed later in this section) makes signal alignment much simpler and more accurate.

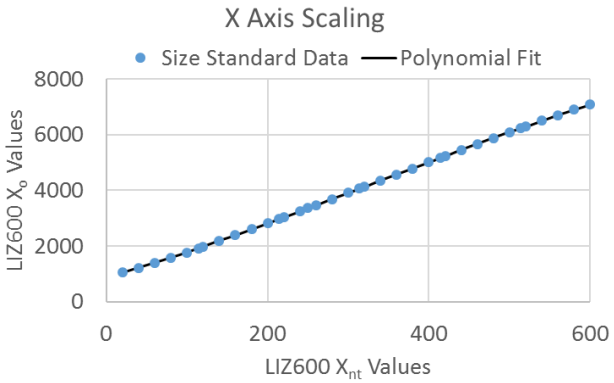
The sequence-alignment step in QuShape performs three processing steps simultaneously, peak picking, sequence alignment, and peak linking. This step frequently does not pick peaks correctly leading to errors in sequence alignment and peak linking, and requiring a very time-consuming manual override. Improvements to the peak picking process (discussed later in this section) alleviate the need for much of the manual override in this step. Finally, when manually overriding calculations, the user interface of QuShape is very difficult to control, as it requires the user to drag and drop assignments in the signal and sequence alignment steps as opposed to entering a corrected numerical value. This often leads to inaccuracies and increased analysis time. The manual overrides built into our analysis method are completely numerical. All of the processing is performed in Microsoft Excel™, which significantly improves the efficiency and control of these necessary overrides.

The main aspect of FAST that needs to be addressed is the lack of calculation transparency, as it depends on Peak Scanner (Applied Biosciences) for much of the data processing. Five out nine calculations in FAST depend on Peak Scanner. However, it is possible for errors to occur during these steps, and Peak Scanner does not have the calculation transparency for the user to see and correct them. To address this, the data after each calculation of our analysis method is completely visible, and many of the parameters for each of the calculations are adjustable.

The first step in the improved analysis method is signal alignment. However, how this analysis step is accomplished is dependent on the correct experimental setup. We use a three capillary experiment, meaning the (+) reaction, (-) reaction, and sequencing are all run in different capillaries with the same primers labeled with the same dye. Included in each of the three capillaries are size standards (DNA oligonucleotides of known length) that allow the signals to be aligned to one another. Setting up the experiment in this way eliminates the need for mobility shift correction, and provides a much simpler and less error prone mechanism for capillary alignment. Mathematically, the capillaries are aligned by scaling the X-axis of each (X_o) to a nucleotide based X-axis (X_{nt}) by fitting a polynomial to the size standard data. This can then predict the relationship between X_o and X_{nt} in all of the capillaries (Figure 7). After the capillaries are aligned, the analysis window can be chosen by visually analyzing the quality of the data and choosing the upper and lower nucleotide bounds.

Size Standard Data

X_{nt}	X_o
20	1067
40	1202
60	1377
80	1570
100	1769
114	1910
120	1973
140	2185
160	2392
180	2605
...	...
600	7087



Polynomial Fit: $X_o = M_3 X_{nt}{}^3 + M_2 X_{nt}{}^2 + M_1 X_{nt} + B$

Figure 7: A cubic polynomial can be fit to the relationship between the migration time and the length of each size standard. X_o is the migration time of the size standard and X_{nt} is its length in nt. M_3, M_2, M_1, and B are constants that can be calculated from the fit. This figure shows how the size standards of one capillary can be scaled to the nucleotide based x-axis. This is used to scale the x-axes of the (+) reaction, (-) reaction, and sequencing traces.

After the capillaries are aligned, the data are subjected to three preprocessing steps prior to analysis: smoothing, baseline correction, and signal decay correction. Smoothing is performed using a moving triangular average method with user defined window size as shown in equations 1 and 2:

$$(\mathbf{1})\ S_i = \frac{R_{i-1} + 2R_i + R_{i+1}}{4}$$

(Window Size=1)

$$(\mathbf{2})\ S_i = \frac{R_{i-2} + 2R_{i-1} + 3R_i + 2R_{i+1} + R_{i+2}}{9}$$

(Window Size=2)

…

S_i is the smoothed intensity at data point i, and R_i is the raw intensity at data point i. This eliminates high-frequency noise in the data (O'Haver, 2016) by performing a weighted average of each raw data point with an adjustable number of the surrounding data points (window size). Any window size can be chosen, but window sizes of greater than 2 often result in a significant loss of information. Although smoothing of any kind results in a small loss of information, it improves the effectiveness of peak picking.

Baseline correction is accomplished by subtracting the minimum number within a surrounding window from a given data point as shown in equation 3:

$$(\mathbf{3})\ B_i = S_i - min[S_{i-w}, S_{i+w}]$$

B_i is the baseline corrected intensity at data point i, S_i is the smoothed intensity from the previous step, and w represents the window size of the baseline correction. The primary reason for baseline correction is to eliminate baseline offset that occurs close to the primer extension initiation (Karabiber et al., 2013). This also results in a decrease in the overlap between peaks making peak picking and Gaussian fitting much simpler.

The final preprocessing calculation is signal decay correction. Reverse transcriptase is not completely processive, so there is a certain probability that it will spontaneously terminate at any nucleotide. Therefore, a smaller population of reverse transcriptase reaches further nucleotides than closer ones causing the intensities to decay quasi-exponentially for longer fragments (Aviran et al., 2011).

Furthermore, the decay is not constant as it depends on the population of reverse transcriptase that stops at any given point. To correct for this phenomena, we divide the intensity of every nucleotide by the sum of the intensities of all the nucleotides after nucleotide i. This is the correction equation that QuShape employs, and can be seen in equation 4 and 5 (Karabiber et al., 2013):

$$(\mathbf{4})\ P_i = \frac{B_i}{E_{last} + \sum_{i+1}^{k} B_{i+1}}$$

E_{last} Criteria:

$$(\mathbf{5})\ \sum_{i}^{k/2} P_i - \sum_{1+k/2}^{k} P_i \cong 0$$

P_i is the probability that reverse transcriptase will terminate at nucleotide i. B_i is the baseline corrected intensity from the previous step, and k represents the last nucleotide within our analysis window. Finally, E_{last} is the estimated sum of the intensities beyond nucleotide k. This estimation is calculated by programmatically varying the E_{last} value to minimize the difference between the sum of the P_i values for the first half of the trace, and the sum of the P_i values for the second half. However, to use this equation before peak area approximation (both our method, and QuShape employ it this way), it must be assumed that applying the formula at every data point is comparable to applying it at every nucleotide.

The peaks that represent individual nucleotides of the RNA molecule must now be identified in the preprocessed data and this is the point that requires the most user involvement. The majority of the peaks are picked by an automated algorithm. This algorithm temporarily plots enhanced data that is the preprocessed data with its second derivative subtracted from it. This emphasizes the peaks tremendously, but leaves the x values the same (Figure 8A). Furthermore, this helps to deconvolute peaks that appear as a shoulder of another peak (Figure 8B) making picking the peaks much clearer. Local maxima are identified as peaks if both of the following criteria are true: $E_X > E_{X-1} > E_{X-2} > E_{X-3} > E_{X-4} > E_{X-5}$ and $E_X > E_{X+1} > E_{X+2} > E_{X+3} > E_{X+4} > E_{X+5}$; where E_X is the enhanced data at a specific point on the X-axis. The enhanced data is only used for peak identification. The X-axis values are correlated with the corresponding Y-axis value

from the preprocessed data for further analysis. After the peaks are picked they must be related to the sequence of the RNA molecule. Sequence assignment is performed using nucleotide number rather than the letters of the sequence for simplicity of calculation. Since the sequencing and the reactions are scaled to the same X-axis, sequence alignment can be performed by assigning a single peak in the sequencing to a single peak in the reactions. All of the surrounding peaks are then assigned based on this one peak. The limitation to assigning the sequence using this method is that the peak picking algorithm is not perfect for all data sets and can miss assign peaks. Missing or extra peaks during peak picking will result in a cascade of nucleotides being one or more positions out of alignment. For this reason, a peak picking check function was incorporated into the analysis tool to tell the user where peaks are likely to be incorrectly assigned. Furthermore, easy-to-use, numerically based manual overrides were incorporated to facilitate this process. Data from multiple experiments were compared to ensure that the peak-picking algorithm was consistent between multiple replicates.

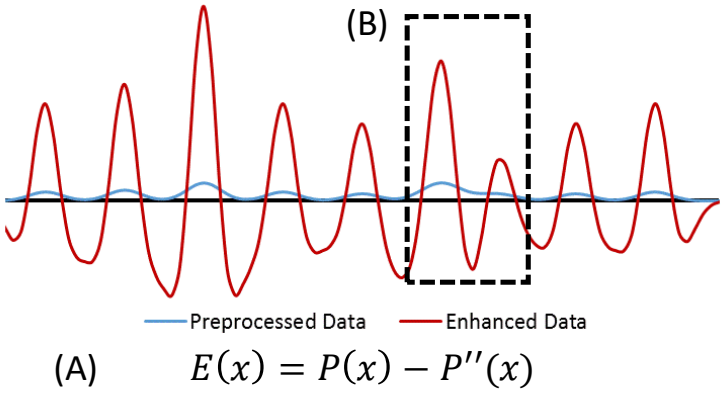


Figure 8: (A) Subtracting the second derivative from the preprocessed data enhances the peaks while leaving the x values the same. (B) The black-dashed box shows how this method improves the picking of shoulders. E(x) is the enhanced data, P(x) is the preprocessed data, and P''(x) is the second derivative of the preprocessed data.

The fourth and final step to processing CE data is reactivity calculation, which involves three computations: peak area determination, scaling, and normalization. The peak areas are calculated using Gaussian deconvolution because there is often overlap between CE peaks. However, if done inefficiently this process can require significant computational resources or a long calculation time. It was determined that the most computationally-efficient method to fit Gaussians to the hundreds of peaks in CE traces is to use a moving optimization window that includes the peak that is being fit, and peaks that are

immediately before and after it. There are three Gaussian parameters that are unique to each peak: position, sigma, and amplitude. To find the best fit for a single peak, trial values for one of the three parameters are tested, and the trial values that give the lowest error when compared to the preprocessed data are chosen (Figure 9). The window then moves one peak to the right, and repeats this process. This is done in three waves of moving optimization: first to optimize the X-axis position of each peak, then the sigma values, and then the amplitude. Once the peak areas are calculated, the (+) and (-) traces can be scaled to one another to correct for load error. This is performed under the assumption that low-area peaks are indicative of reverse transcriptase spontaneous termination (i.e. we would expect them to be equivalent in area in both traces). A scaling factor, α, can be calculated by dividing the average of the lowest 20% of (+) peaks by the average of the lowest 20% of (-) peaks. The minus peaks are then scaled to the plus by multiplying them by α. Finally, the normalized reactivity values for each nucleotide can be calculated. First, the background is removed by subtracting the (-) peak areas from the corresponding (+) peak areas. Background subtracted areas are then normalized by dividing by the average of the top 10% of peak areas. Outliers are excluded when determining the

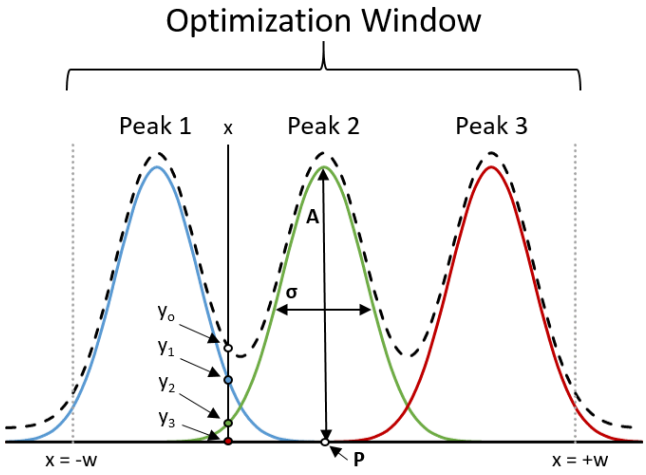


Figure 9: Gaussian fitting of Peak 2. P is the position of the peak, σ is the width at half the height of the peak, and A is the amplitude of the Gaussian function. y_o is the preprocessed intensity value of the unfit data at point x, and y_1, y_2, and y_3 are the intensity values of the Gaussian fits of peaks 1, 2, and 3 respectively at the same x value. This figure shows how to calculate the Fit_{error} for a trial P, σ, or A value for a single this peak. Many trial values are tested for each of these three values, and the trial values for each that give the lowest Fit_{error} are elected to represent the Gaussian of this peak.

top 10% if they are greater than the first quartile plus 1.5 times the interquartile range. However, no more than 10% of peaks can be excluded. Normalizing in this way allows a value of 1.0 to represent the average value of a highly reactive nucleotide.

Once reactivity values are determined, different replicates can be used to compare and correct for missing or extra nucleotides in one or more of the replicates. These data can then be used as constraints in RNA structure prediction, or in the case of protection experiments, analyzed for the footprint of the RNA binding molecule.

Conclusions and Future Directions

The process of optimizing RNA probing is well underway, the CE data analysis improvements have allowed us to predict the structure of the HTLV-1 5'UTR, a very long and complex RNA. However, there is still much that remains to be done as far as optimization and analysis are concerned. Primer extension needs to be further optimized to allow for the efficient analysis of HIV-1 RNA probing data. This will allow probing of the HIV-1 5'UTR-LysRS interaction footprint. Additionally, it will then be possible to collect HIV-1 5'UTR SHAPE data that can be used as a validation of our new analysis method.

There are also many improvements that we still would like to make to the probing analysis tool to facilitate CE data analysis. While none of the planned improvements will affect the current signal processing methods, they improve the user experience. First, we would like to incorporate the ability to convert FSA files (the file format in which CE data is packaged) into text files that can be opened and easily manipulated. It would then be easy to program the analysis tool to copy the data from the raw file into the analysis tool automatically. Currently we use an outside program to convert the files and copy and paste the data manually. We would also like to incorporate a size standard picking mechanism into the analysis tool. Currently, Peak Scanner (Applied Biosystems) is employed to find the X-axis values of the size standard peaks, which are then copied into the program manually. Thus, incorporating data

importation and size standard picking in these ways would eliminate the involvement of any external programs in CE data analysis. These improvements are currently being implemented.

Probing optimization and analysis improvements will hopefully help future members of our group and others study RNA structure and function through chemical and enzymatic probing. The crowning achievement of probing experiments in our lab following the improved probing analysis method has been determining the HTLV-1 5'UTR secondary structure using SHAPE (data discussed in Chapter 3). This has contributed significantly to the study of this virus in our lab and pioneered the way for further SHAPE experimentation. After optimization of SHAPE on the HIV-1 5'UTR is complete, resolving this well-studied secondary structure will be used to validate our probing analysis method.

Acknowledgements

I would like to thank Dr. William Cantara directing the probing data analysis methodology project, and for allowing me to take the lead in prototyping certain programmatic components myself. I would also like to thank him for mentoring me through the process of learning all of these complex techniques. Finally, I would like to thank Weixin Wu for partnering with us in this project and for pioneering SHAPE probing in our lab with the HTLV-1 5'UTR.

Chapter 2: Human Immunodeficiency Virus Type I

Introduction

HIV-1 is well known as the cause of acquired immunodeficiency syndrome (AIDS), which is characterized by the depletion of CD4+ helper T-cells resulting in opportunistic infections or cancers that ultimately lead to the death of the patient (Gallo et al., 2003). HIV-1 can be transmitted by blood-to-blood contact with an infected person (Drucker et al., 2001), vertically from mother to child either *in utero* or through breast feeding (Peckham et al., 1995), or sexually in both heterosexual and homosexual relationships. The 2014 World Health Organization (WHO) statistics estimate that 37 million people are infected with HIV, a global prevalence of 0.8% (4.5% in Africa). Furthermore, they estimate that more than 34 million people have died of AIDS-related causes since the beginning of the epidemic (World Health Organization, 2015). HIV-1 belongs to the genus lentivirus, which includes simian immunodeficiency virus (SIV), as well as other mammalian immunodeficiency viruses, all of which are characterized by their slow progression to disease (Freed et al., 2007). With production of antiretroviral medications (ARVs), this long latency period can be exploited to diagnose and treat HIV-1 before it progresses to AIDS. As of 2012 there were 24 Food and Drug Administration approved ARVs that target 5 different viral mechanisms: RT, integration, protease cleavage, membrane fusion, and receptor binding (Arts et al, 2012). However, HIV-1 is a highly mutagenic virus with the capacity to develop resistance to many different ARVs. Therefore, although there are many drugs available, the pressure to find a cure and to combat resistance emphasizes the importance for the scientific community to continually identify molecular mechanisms that could eventually be used as drug targets.

One potential target is the RNA genome itself. While primarily responsible for carrying the genetic material of the virus, the RNA genome also serves to regulate various functions of the viral lifecycle. Many of these regulating elements are located in the highly-structured 5'UTR of the viral RNA. The TAR and polyA stem loops, located at the 5' end of the 5'UTR, are involved in genomic circularization and minus-strand transfer during RT (reviewed in Basu et al., 2008). Additionally, these sequences serve as the

transcriptional stimulator for viral RNAs in the integrated proviral DNA (Bannwarth et al., 2005). Downstream from these stem loops are the TLE stem loop and the PBS, which play roles in RT regulation by serving as the location of primer binding (Mak et al., 1997). It has also been proposed that the TLE facilitates RT by localizing the primer (Jones et al., 2013; Jones & Cantara et al., 2014). At the very 3' end of the 5'UTR are the Psi stem loops which participate in viral packaging and genomic dimerization (Lu et al., 2011; Skripkin et al., 1994). The concentration and diversity of functional elements in this highly-conserved segment of the HIV-1 viral RNA makes the 5'UTR an ideal candidate for further scientific study.

Our primary interest with the 5'UTR is the localization of the tRNA primer through the interaction between lysyl-tRNA synthetase (LysRS) and the TLE. As mentioned earlier, many retroviruses hijack host cell tRNAs to prime RT and the primer for HIV-1 RT is $tRNA^{Lys3}$. Both $tRNA^{Lys3}$ and LysRS are selectively packaged into HIV-1 virions, which originally indicated the synthetase's involvement in selectively-packaging the tRNA primer (Cen et al., 2001). Furthermore, other studies have shown that knocking down LysRS in cells results in decreased primer incorporation into virions and a less infectious virus (Guo et al, 2003) whereas overexpression of LysRS had the opposite effect (Guo et al. 2005). Through analysis of the secondary structure of the 5'UTR, it can be seen that the loop of the TLE hairpin has high sequence similarity to the anticodon loop of $tRNA^{Lys3}$, and contains many of the identity elements for specific recognition by LysRS. Following up on this finding, it was shown that this element clearly mimics $tRNA^{Lys3}$ as it successfully competes with the tRNA for binding with LysRS (Jones et al., 2013). Furthermore, mutation of the UUU in the loop to CCC caused approximately a 40% reduction in viral infectivity (Jones et al., 2013). Molecular envelopes of the PBS/TLE construct and $tRNA^{Lys3}$ derived from small-angle X-ray scattering (SAXS) were highly homologous, demonstrating that the TLE resembled the tRNA in 3D structure as well as sequence (Jones & Cantara et al., 2014). Finally, it has been demonstrated that LysRS binds to the PBS/TLE with the first stem loop of Psi (SL1), better than to the PBS/TLE alone, and addition of SL2 improves binding further (Jones et al., 2013). As neither SL1 or SL2 resemble tRNAs, this trend

could suggest that HIV-1 has evolved structures beyond the PBS/TLE domain to facilitate the binding of LysRS and the localization of the tRNA primer. The first section of this chapter will discuss RNA probing of the interaction between LysRS and the HIV-1 5'UTR using RNase protection and SHAPE analyzed by CE to map the HIV-1 genomic RNA-LysRS interaction.

We hypothesized that the structure of the HIV-1 5'UTR as a whole will yield insights into its function. Many studies using various techniques have attempted to determine the secondary and tertiary structure of the HIV-1 5'UTR. However, to date no one has solved its complete 3D structure. The secondary structure of the 5'UTR and the rest of the HIV-1 genome has been solved using SHAPE. However, the 5'UTR, particularly Psi hairpins, are believed to adopt alternative secondary structures depending on the stage in the lifecycle. Additionally, 3D models have been proposed for large portions of the 5'UTR using SAXS and nuclear magnetic resonance spectroscopy (NMR). Along with revealing the structural similarities between the HIV-1 PBS/TLE and $tRNA^{Lys3}$, the same study modeled the 3D structures of two other segments of the HIV-1 5'UTR based on SAXS data: TAR/polyA and Psi (Jones & Cantara et al., 2014). Furthermore, a larger structure has been solved of the minimal RNA required for genomic RNA packaging by NMR. This RNA consists of the Psi and U5 region of the 5'UTR, deleting the TAR and polyA hairpins, and replacing the PBS/TLE and the DIS with GAGA tetraloops (Keane et al., 2015). The second section of this chapter will discuss how SAXS has been used to build upon this NMR structure. Additionally, we will show how a comparison between structures from subgroup A and B imply the conservation of the TLE mechanism between these two viral subgroups.

Materials and Methods

RNA preparation

RNA probing experiments were performed with variants of the HIV-1 5'UTR(356), which consists of the first 356 nucleotides of the HIV-1 viral RNA. For preparation of HIV-1 5'UTR(356) constructs, see "RNA preparation" in Chapter 1.

All of the RNAs studied by SAXS were variants of the HIV-1 5'UTR ΔTAR/polyA, which will be referred to as the HIV-1 5'UTR(240). Mutations such as extensions and deletions to this construct were made to the HIV-1 5'UTR(356) plasmid. Transcription templates were obtained by PCR from plasmids containing the HIV-1 5'UTR(356) sequence. RNAs were *in vitro* transcribed and purified in the same manner described for other RNAs (see Chapter 1). RNAs were folded in 50 mM HEPES (pH 7.4) buffer by heating to 80°C for 2 min, cooling to 60°C for 2 min, incubating at 37°C for 30 min, and cooling on ice for at least 30 min. The folded RNAs were then purified by size exclusion chromatography, and stored at 4°C until ready for analysis.

Protein preparation

Two different constructs of human LysRS were used in this study: LysRS triple mutant (LysRS-3M), and LysRS-ΔN65. LysRS-3M is a primarily monomeric construct obtained by substituting three amino acids of the dimerization interface (R246, E265, and F283) with alanine residues (Kovaleski et al., 2006). This mutant was incorporated into this study with the hopes of facilitating SAXS studies by alleviating oligomerization of LysRS, which leads to sample heterogeneity. The second construct, LysRS-ΔN65, has the first 65 N-terminal amino acids deleted. This is useful as it is known that these 65 residues are unstructured, which may lead to conformational heterogeneity, and the deleted residues are not necessary for tRNA binding, or aminoacylation (Shiba et al., 1997). It is also the construct that was used in most of the *in vitro* studies characterizing this interaction.

Detailed protocols for the purification of LysRS-3M and LysRS-ΔN65 can be found in Kovaleski et al. (2006) and Jones et al. (2013), respectively. BL21DE3 cells containing plasmids encoding LysRS-3M and LysRS-ΔN65 genes were grown in liquid media containing 100 μg/μL ampicillin, and induced by the addition of 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were then lysed using sonication, and nucleic acids were precipitated using polyethyleneimine (PEI) precipitation. The protein was then precipitated with ammonium sulfate precipitation, resuspended, and purified using Ni affinity

purification. It was then buffer exchanged into 2X storage buffer (80 mM HEPES, 300 mM NaCl, 4 mM DTT), and diluted with 80% glycerol to achieve 40% glycerol final storage solution. The concentration was determined via Bradford assay.

Small-angle X-ray scattering

Size exclusion chromatography (SEC) purified RNAs (Jones & Cantara et al., 2013) were sent to the Advanced Light Source beamline 12.3.1 at Lawrence Berkeley National Labs for SAXS data collection. Data were then analyzed using the ATSAS package (Konarev et al., 2006). PRIMUS was used for data preprocessing (Konarev et al., 2006) and envelopes were generated using DAMMIN (Konarev et al., 2006).

Ribonuclease/SHAPE protection assays

To probe the interaction between the HIV-1 5'UTR(356) ΔDIS A34U RNA and LysRS-3M, RNase protection assays were used. To probe the interaction between the HIV-1 5'UTR(356) ΔDIS A34U RNA and LysRS-ΔN65 SHAPE protection assays were used. These assays consisted of five reactions: background (no probing reagent), reaction with probing reagent without protein, and three reactions with probing reagent with pre-bound serially diluted protein. Reactions were initiated with the addition of buffer to the background and probing reagent to the other four reactions. The reactions were incubated for the optimized amount of time. The RNase protection assays were quenched by ethanol precipitating the reactions with premixed precipitation/inactivation buffer (Ambion), and the SHAPE protection assays were quenched by ethanol precipitating with glycogen. The reactions were then reverse transcribed by annealing NED-labeled primers and following the Superscript III primer extension protocol.

Sequencing

For sequencing information, see "Sequencing" in Chapter 1

Results and Discussion

Section I: LysRS interaction with HIV-1 5'UTR(356) probed by RNases and SHAPE

The HIV-1 viral genome is more than 9 thousand nt long, and yet the PBS is only 18 of those nucleotides. Somehow, the tRNA primer needs to be shuttled to the vicinity of the PBS so it can be efficiently annealed for RT to initiate. Logically, this localization would be facilitated by a protein that binds to $tRNA^{Lys3}$, the primer for HIV-1 RT. As previously mentioned, LysRS is packaged into HIV-1 virions to facilitate primer packaging (Cen et al., 2001), and $tRNA^{Lys3}$ can be competed off by the TLE located near the PBS in the 5'UTR (Jones et al., 2013). Additionally, 3D modeling based on SAXS data demonstrated that the TLE strongly mimics the three dimensional structure of a tRNA (Jones & Cantara, 2014). These results have led our lab to propose that LysRS is central to the mechanism of localizing the RT primer. We hypothesize that LysRS shuttles the tRNA primer to the PBS/TLE domain of the 5'UTR, after which the

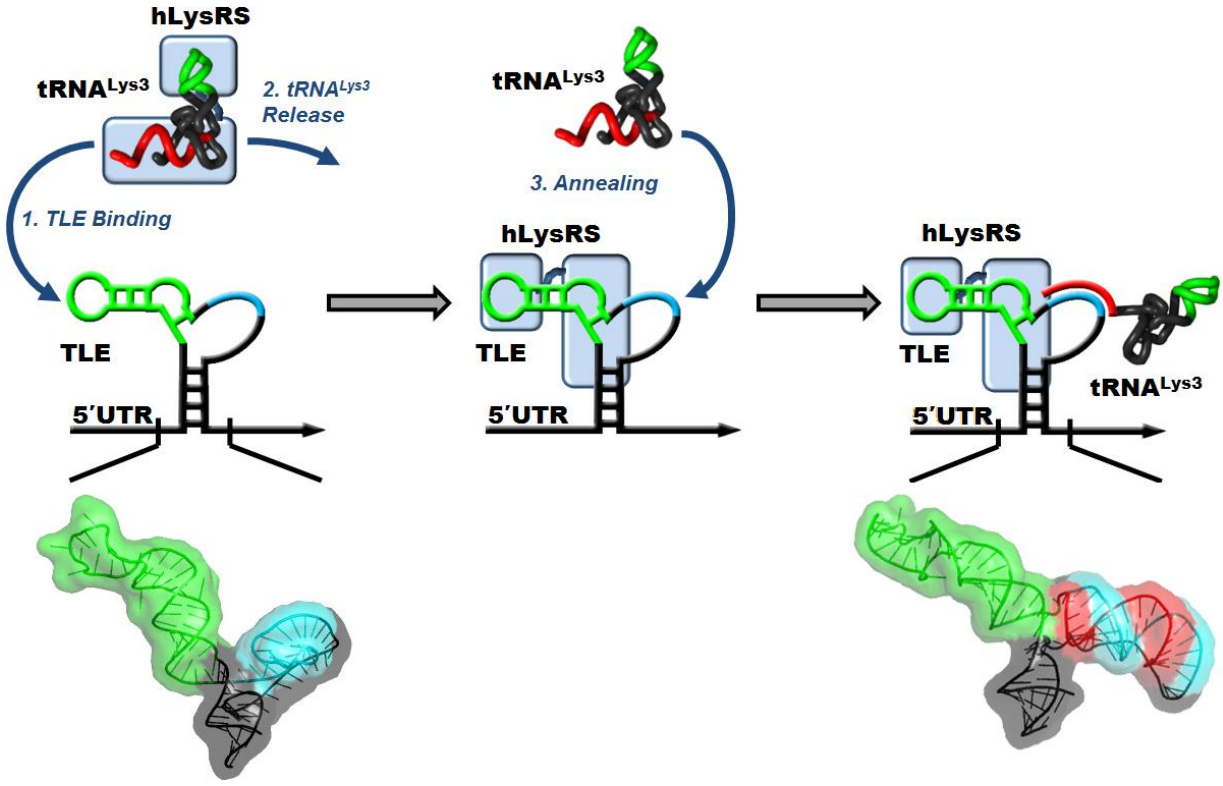


Figure 10: The proposed mechanism for $tRNA^{Lys3}$ localization involves LysRS. LysRS chaperones the tRNA to the 5'UTR after which it binds to the TLE, releasing the tRNA in close proximity to the PBS; thus, facilitating primer annealing via localization. The SAXS-derived models below are of the PBS/TLE RNA with (right) and without (left) an 18 nt primer annealed (Jones & Cantara et al., 2014). (Figure prepared by William Cantara.)

RNA	hLysRS dN65 K_d
$tRNA^{Lys3}$	407 ± 33
PBS-TLE	383 ± 18
PBS + Psi SL1	281 ± 25
PBS + Psi SL1, SL2	85.4 ± 5.4

Figure 11: K_d values of LysRS-ΔN65 with PBS/TLE and $tRNA^{Lys3}$ are very similar. However, as the construct is extended to include stem loops of the Psi region the K_d values drop approximately four-fold (Jones et al., 2013).

tRNA is competed off by the TLE, localizing the primer for annealing prior to RT (Figure 10). It has also been shown that adding stem loops from Psi to the PBS/TLE RNA significantly decreases the K_d (Figure 11) (Jones et al., 2013). This suggests involvement of these stem loops in the specificity of LysRS binding, leading to the hypothesis that HIV-1 has developed structures beyond tRNA mimicry to enhance competitive binding with $tRNA^{Lys3}$ for LysRS. This section will explore how RNA probing is being used to interrogate the interaction between LysRS and the HIV-1 genome, and inform this hypothesis.

The first probing technique employed to investigate this interaction was RNase digestion with RNases T1 and V1 analyzed by CE. Qualitative analysis of these data suggested that LysRS-3M was binding to the TLE and to SL1 and SL2 of the HIV-1 5'UTR ΔDIS A34U RNA (Figure 12). Unfortunately, it was found that there was strong signal overlap between channels of the same capillary rendering quantitative analysis with the new method described in Chapter 1 impossible. Prior to completion of the analysis tool, the amount of primer extension product optimal for analysis had not been determined, so signal overlap was the result of overloading the capillary with sample. These experiments will be repeated using new, optimized conditions with the newly-developed quantitative analysis methods.

As mentioned in Chapter 1, SHAPE is also being optimized for experiments on the HIV-1 5'UTR. However, for both structure probing and protein protection experiments the same road block has been preventing progress. That is, large peaks in the minus traces indicative of low processivity of reverse transcriptase is resulting in unanalyzable data. There are two likely issues that could be causing reverse transcriptase to fall off, and both are in the process of being tested. The first is the simple explanation: RNA degradation. Degraded fragments of RNA at varying lengths would make it appear as if reverse transcriptase were falling off prematurely, when in actuality, the polymerase ran out of template at the

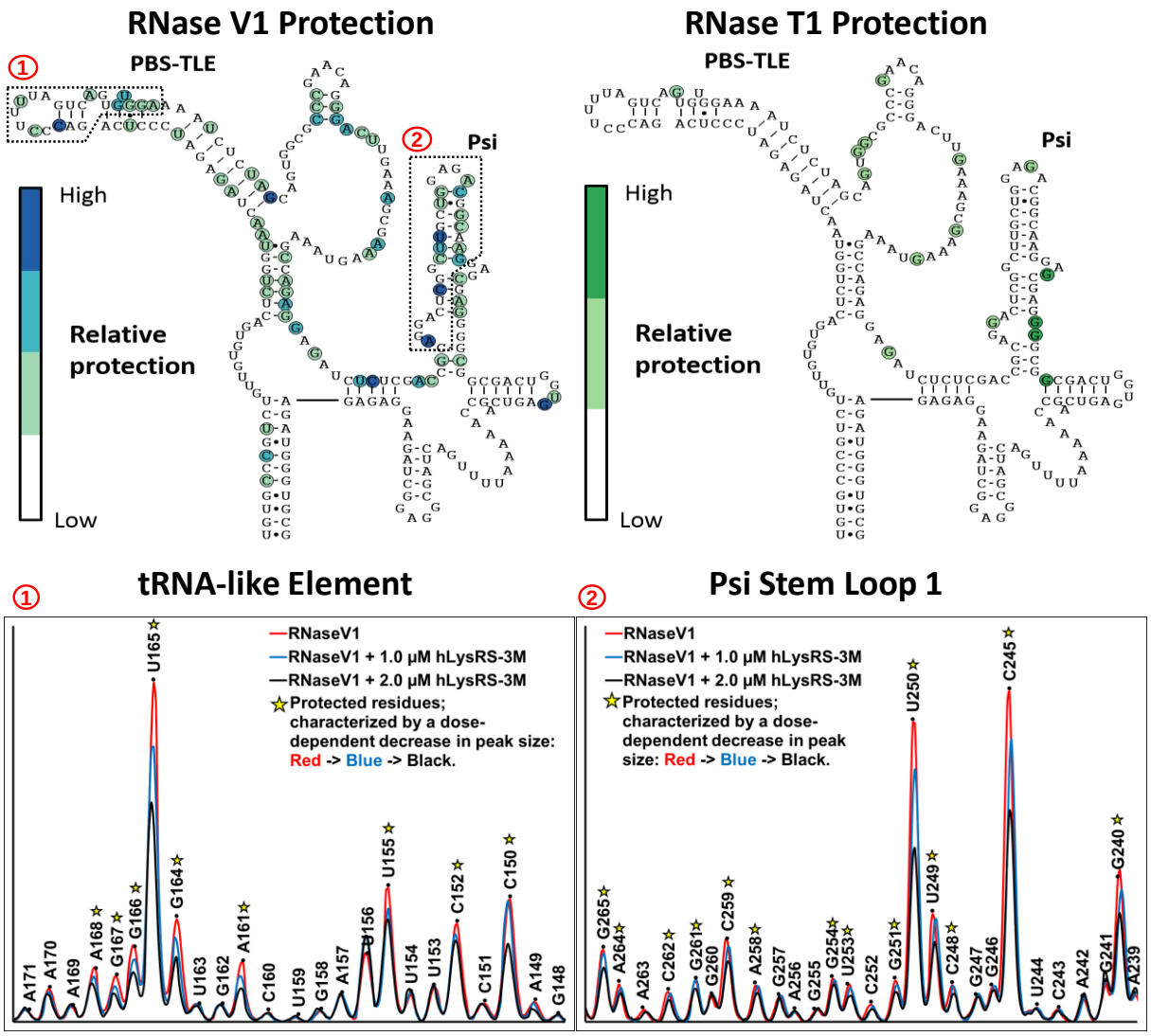


Figure 12: RNase V1 and T1 data plotted onto part of the secondary structures of the HIV-1 5'UTR ΔDIS A34U. Decrease in peak area can be clearly seen in the images of the electropherograms shown of the TLE and SL1 as LysRS is titrated in. However, this data has not been quantitatively analyzed.

end of a degraded fragment. Alternatively, the HIV-1 5'UTR is more highly structured than most RNAs, so it is possible that the thermal denaturing cycle used prior to primer extension reactions may not fully denature the RNA. In this case, secondary and tertiary structural elements could be interfering with polymerase processivity. To overcome this, fresh RNA is being prepared and alternative thermal denaturing conditions will be tested.

Section II: SAXS study of HIV-1 5'UTR ΔTAR/polyA

The 5'UTR of HIV-1 is a highly-conserved, heavily-structured RNA that is involved in many functions of the retroviral lifecycle. Although many studies have proposed models and structures of small pieces of the 5'UTR, no one has reported a structure or model of it in its entirety. Our lab alone has proposed 3D models based on SAXS data of the TAR/polyA stem loops, the Psi stem loops, and the PBS/TLE region with and without primer annealed (Jones & Cantara et. al., 2014). These structures also served to reveal the 3D similarities between the TLE and $tRNA^{Lys3}$ (Jones & Cantara et. al., 2014). In this section, a recent NMR structure of the HIV-1 minimal packaging element will be discussed. It is currently the largest RNA structure to be experimentally determined by NMR (Keane et al., 2015). In this section, the use of SAXS to expand on this structural ground work, and attempts to model even larger constructs of the 5'UTR will also be discussed.

The goal of SAXS is to generate electron pair-distance distribution functions (PDDF) from which low resolution envelopes of molecules can be calculated. The PDDF is a histogram of all inter-electron distances in a molecule. We were able to use this method to calculate the envelope of the 5'UTR(240) ΔDIS (Figure 14A). However, the resolution of SAXS is not sufficient for unambiguous assignment of one region of the RNA from another within an envelope. For this reason, a series of extension and deletion mutants were made to this RNA construct to attempt to determine how the RNA fits into the envelope (Figure 13). Although many of these mutants resulted in good quality data, due to limitations of the DAMMIN algorithm, their resulting envelopes did not converge to a unique shape. The 5'UTR(240) ΔDIS ΔPBS/TLE (which is very similar to the construct

Name	Length (nt)	Prepared by:
5'UTR(240) WT	240	J. Hatterschide
5'UTR(240) ΔDIS	235	J. Hatterschide/W. Cantara
5'UTR(240) ΔDIS ΔPBS/TLE	154	J. Hatterschide/W. Cantara
5'UTR(240) ΔDIS TLE extension	255	J. Hatterschide
5'UTR(240) ΔDIS ΔSL1	217	W. Cantara
5'UTR(240) ΔDIS SL1 extension	255	J. Hatterschide/W. Cantara
5'UTR(240) ΔDIS GU120AC	235	J. Hatterschide/W. Cantara
5'UTR(240) ΔDIS U288C G294A	235	J. Hatterschide

Figure 13: This table includes all of the HIV-1 5'UTR(240) RNA constructs that were prepared for this SAXS study, their length in nt, and the person(s) primarily responsible for their mutagenesis and preparation.

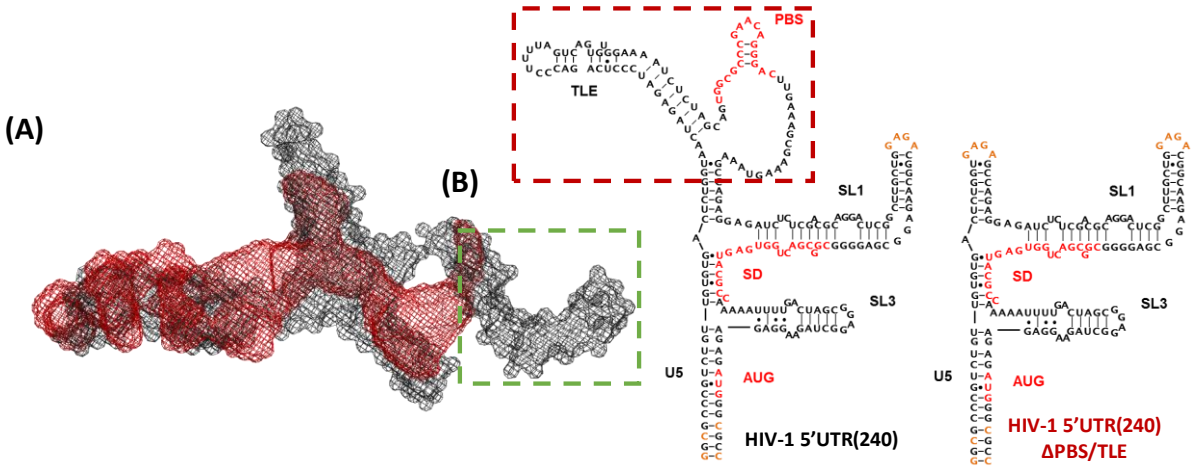


Figure 14: (A) The envelope of the 5'UTR(240) ΔDIS can be seen in the black mesh with the envelope from the 5'UTR(240) ΔDIS ΔPBS/TLE overlaid in red. The secondary structure of the UTR(240) ΔDIS and UTR(240) ΔDIS ΔPBS/TLE can be seen in to the right. The UTR(240) ΔDIS ΔPBS/TLE is the previous construct with the region in the red-dashed box mutated to a GAGA tetraloop. (C) The extra space remaining in the UTR(240) ΔDIS envelope, green-dashed box, resembles the L-shape of a tRNA which is the shape that the TLE has been established to adopt. This gives us a working model of the identities of each lobe of the envelope. (Figure modified from William Cantara)

solved by Keane et. al.), was one mutant that yielded a useful envelope (Figure 14A). When overlaid with the envelope of 5'UTR(240) ΔDIS, it can be seen that additional density looks remarkably like a tRNA. This fits with previous data showing the TLE mimics the 3D structure of a tRNA, suggesting that this section of the SAXS envelope corresponds to the PBS/TLE region (Figure 14B). However, in order to assign the lobes more rigorously, models of the ΔDIS and extension and deletion constructs are currently being generated (W. Cantara, unpublished data). From these, theoretical PDDFs can be calculated and compared to the experimental data. However, the modeling required for an RNA of this size will be time-consuming and new methods will need to be developed in order to yield reliable models.

Using preliminary helix assignments, it was possible to analyze the model of the UTR(240) ΔDIS in comparison to the structure determined by Keane et al. It was found that our envelope of this very similar construct agrees with the NMR model in terms of helical dimensions. However, the interhelical orientations of the NMR structure are very different from those in the SAXS envelope. When the U5-AUG duplex, and the PBS/TLE hairpin of the all atom NMR model are aligned with the corresponding portions of the SAXS envelope, the SL1 and SL3 hairpins of the model cannot align with other sections of the

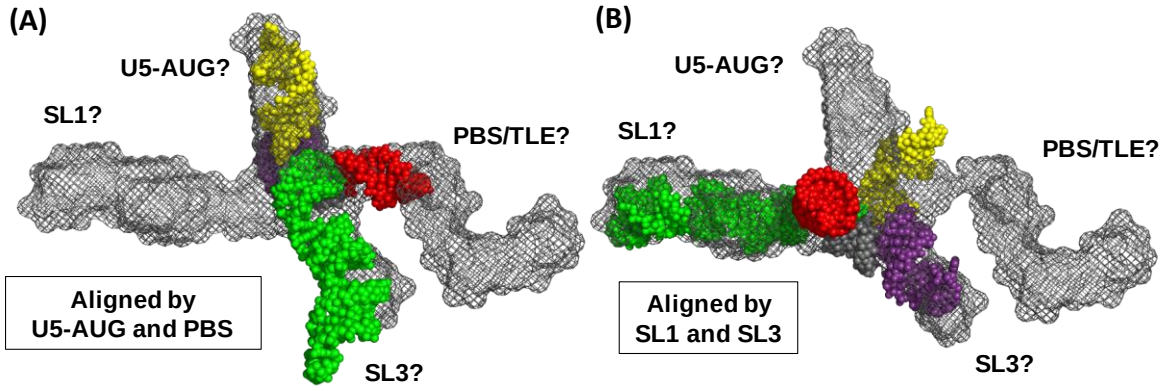


Figure 15: The UTR(240) ΔDIS envelope has similar volume and dimensions to the Keane el al. model. The predicted lobes of the SAXS envelope are labeled in black, and the helices of the model are color coded: yellow is the U5-AUG, red is the PBS/TLE truncated hairpin, green is SL1, and purple is SL3. (A) The UTR(240) ΔDIS SAXS envelope aligned with the Keane et al. all atom model envelope by the U5-AUG duplex and the PBS/TLE, cannot accommodate SL1 and SL3 of the model. (B) The same problem is encountered when the model and the envelope are aligned by the other two helices. (Figure modified from William Cantara).

envelope (Figure 15A). Moreover, when SL1 and SL3 are aligned with their predicted lobes of the SAXS envelope, PBS/TLE and U5-AUG do not match the envelope (Figure 15B). Since the NMR study did not employ a method for determining global interhelical angles, it is possible that the interhelical orientations of the Keane et al. model could be incorrect. Alternatively, the inclusion of the large polyanionic PBS/TLE domain may alter the interhelical conformation. The SAXS-based modeling should be able to unambiguously distinguish the overall global topology resulting in correct determination of interhelical orientations.

So far, the discussion has been about the 5'UTR(240) that originates from the sequence of the 5'UTR of the NL4-3 isolate. HIV-1 NL4-3 is a circulating recombinant form with a 5'UTR corresponding to group M, subgroup B, the primary subgroup of HIV-1 that affects Europe, the Americas, and Oceania (Hemelaar et al., 2006). However, if the structures and mechanisms that are discovered in the NL4-3 isolate are not conserved among many different subgroups of the virus, then they will likely not contribute to the understanding of important facets of the HIV-1 lifecycle. The SAXS-based envelopes of a region similar to the UTR(240) in the MAL isolate's 5'UTR have been calculated (W. Cantara, unpublished data). The MAL isolate is a circulating recombinant form, the 5'UTR of which corresponds to group M, subgroup

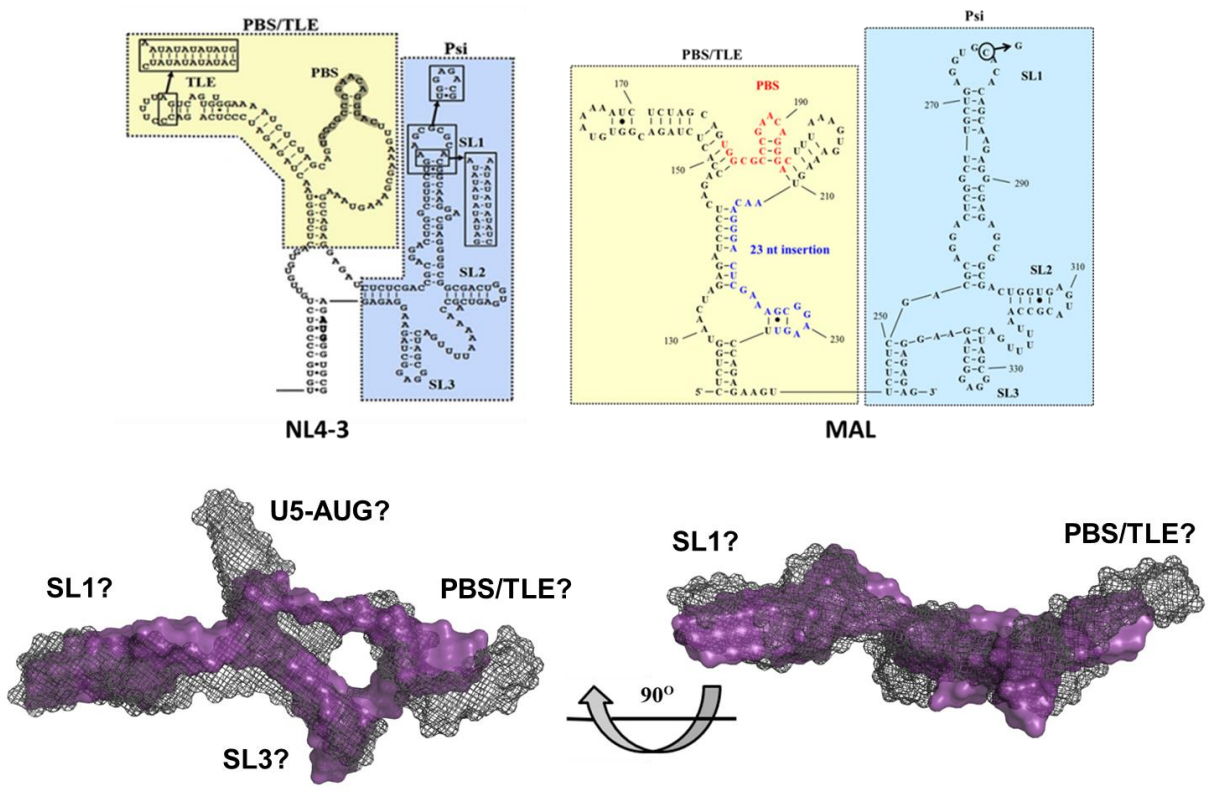


Figure 16: The overlaid NL4-3 and MAL SAXS envelopes display very high structural similarity despite the relatively significant differences in sequence. This structural similarity implies that many of the structural and functional studies perform on the NL4-3 isolate's 5'UTR can be translated to the MAL isolate. (Figure modified from William Cantara).

A (Gao et al., 1998), a subgroup that primarily affects western, central, and eastern Africa, eastern Europe, and central Asia. (Hemelaar et al., 2006). Interestingly, the sequence of MAL contains a 23-nucleotide insertion that alters the secondary structure of this region of the 5'UTR (Goldschmidt et al., 2004). When the NL4-3 and MAL SAXS envelopes are overlaid, it can be seen that despite the differences in sequence and secondary structure, the 3D structures of these two large RNAs are very similar (Figure 16). This conservation in 3D structure implies that many of the mechanisms studied with NL4-3 RNAs will be conserved in MAL and possibly other subgroups as well. Finally, the 3D similarities between these two RNAs explain how LysRS also selectively binds to the PBS/TLE domain of MAL.

Although an all-atom model has not been modeled for the HIV-1 NL4-3 5'UTR(240) ΔDIS, this data has been immensely useful in comparison to other data sets. Not only has it informed our perception of

this complex RNA 3D structure, but it has also contributed to the understanding of more detailed structures such as that of Keane et al. It has also helped to suggest that the 3D structure of the 5'UTR is a feature that is conserved between subgroups of HIV-1.

Conclusions and Future Directions

Despite initial progress, RNA probing experiments with the HIV-1 5' UTR and LysRS are still in the process of being optimized. Direct binding data suggests that the addition of the Psi stem loops SL1 and SL2 enhance LysRS binding (Jones et al., 2013). Initial CE traces have been analyzed qualitatively and suggest that LysRS is binding to both the TLE and the first two Psi stem loops—in agreement with the direct binding data. However, once optimization is complete, experiments can be analyzed quantitatively to provide a more detailed characterization of the interaction between LysRS and the viral genome. These experiments will help to inform the hypothesis that HIV-1 could have evolved more extensive structures than just tRNA mimicry to compete with the tRNA for LysRS binding.

The envelope of the HIV-1 5'UTR(240) ΔDIS has been modeled based on SAXS data. This has been very useful in comparisons with other envelopes and models. When the SAXS envelope of the 5'UTR(240) ΔDIS and 5'UTR(240) ΔDIS ΔPBS/TLE are overlaid we notice that the remaining density has the L-shaped structure of a tRNA. This suggests that the 3D tRNA mimicry that was described in Jones & Cantara et al. is conserved in the context of this larger RNA. The 5'UTR(240) ΔDIS envelope was also able to reveal that the inter-helical orientation of the Keane et al. NMR structure is likely incorrect. Finally, we saw that the 3D structure of the PBS/TLS and Psi regions of the 5'UTR is conserved between subgroup A and B through comparison between the NL4-3 5'UTR(240) and the corresponding region of the 5'UTR from the MAL isolate. An all-atom model will be generated in the future. It should be noted that a model of the HIV-1 5'UTR(240) ΔDIS would be one of the largest experimentally informed models of the HIV-1 5'UTR and will require development of new methods.

The 5'UTR is a very complex RNA, but its involvement in a wide array of viral lifecycle events makes it a primary target for further scientific investigation. Both the RNA probing and SAXS experiments will contribute to a more complete understanding of the structure and function of the HIV-1 5'UTR, and provide the basis for future studies.

Acknowledgements

I would like to thank Dr. William Cantara for doing most of the SAXS data analysis, and creating their figures. I would also like to thank Roopa Comandur and Erik Olson for contributing to the MAL project.

Chapter 3: Human T-cell Leukemia Virus Type I

Introduction

HTLV-1 is a delta-retrovirus known to be associated with multiple different diseases that are found in a small fraction of infected individuals. The most common and severe of these diseases are adult T-cell leukemia/lymphoma (ATL) and HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TPS), but less common diseases include HTLV-1-associated dermatitis, ocular lesions, and inflammatory arthropathy and polymyositis (Lairmore et al., 2012). ATL, which occurs in about 3-5% of infected individuals after a prolonged latency period (reviewed in Fujii et al., 2007), is a highly aggressive form of non-Hodgkin's lymphoma which is clinically identified by clonally expanding CD4+ CD25+ T-cells (Cook et al. 2012). HAM/TSP affects from 0.1 to 2% of individuals infected with HTLV-1, and is characterized by the invasion of the cerebrospinal fluid with CD4+ and CD8+ T-cells, damaging the spinal cord and resulting in spasticity in the lower half of the body (Osame, 2002; Irish et al., 2009).

The transmission of HTLV-1 is similar to that of HIV-1 and other retroviruses, but the efficiencies of each mode of transmission differ (Biggar et al., 2006; Goncolves et al. 2010). HTLV-1 is not a global pandemic like HIV-1, but rather is highly endemic to a few large areas including the Caribbean, Japan, and West and Central Africa. Although these regions of Africa (primarily Nigeria and the Democratic Republic of the Congo) are estimated to be the most endemic with HTLV-1, comparatively few large population studies have been undertaken in these areas which could affect global approximations of the number of HTLV-1 carriers (Gessain et al., 2012). With this in mind, the most recent epidemiological models of HTLV-1 infection predict that there are between 5 and 20 million people infected worldwide (De Thé et al., 1993; Gessain et al., 2012).

Attempts to control and treat ATL and HAM/TSP have been met with limited success (Goncolves et al., 2010) and there is no cure for either of these diseases. One of the only treatments found to result in a high remission rate among chronic ATL patients (other ATL types not tested) is a combination of arsenic trioxide, zidovudine (AZT), and interferon alpha (Kchour et al., 2009). Clinical studies testing the

efficacy of AZT and interferon alpha on more aggressive types of ATL have been met with more moderate remission rates, and all patients eventually relapsed (Matutes et al., 2001). Furthermore, treatment with corticosteroids, interferon alpha, and interferon beta1a have been the only regimens with moderate success in treating HAM/TSP (Goncalves et al., 2010). It has been shown that high proviral load is a precursor to ATL progression (Akbarin et al., 2013), suggesting that treatments resulting in lower proviral load could be useful in preventing HTLV-1 progression ATL. This justifies exploring the HTLV-1 lifecycle through basic scientific research to uncover more virus-specific target mechanisms.

As discussed previously, one potential drug target could be the viral mechanisms of primer packaging and localization. HTLV-1 requires the 3' end of $tRNA^{Pro}$ to prime RT, but currently the mechanism of packaging is still unknown (Seiki et al., 1982). However, it is known that other complex retroviruses such as HIV-1 and Rous sarcoma virus (RSV) package aminoacyl-tRNA synthetases specific to their tRNAs, while simple retroviruses such as Molony murine leukemia virus do not (Cen et al., 2002). As HTLV-1 is also a complex retrovirus, it is possible that it packages glutamyl-prolyl-tRNA synthetase (EPRS) in a similar manner. A recent model produced by our lab of the HTLV-1 5'UTR secondary structure based on SHAPE data revealed a stem loop that resembles the anticodon stem loop of $tRNA^{Pro}$. Much like HIV-1, this putative TLE lies just 5' to the PBS, suggesting that EPRS may be involved in the packaging of the HTLV-1 primer. On the other hand, a recent study has demonstrated that HTLV-1 infection stimulates the production of a 3' fragment of $tRNA^{Pro}$, and that HTLV-1 packages both $tRNA^{Pro}$ and the fragment into virions (Ruggero et al., 2014). They suggest that this fragment could be the primer for RT along with or in place of the full-length tRNA. Although this in itself does not discount the involvement of EPRS, it does suggest that there may be an alternate route of packaging primer RNAs as EPRS would likely not specifically recognize the tRNA fragment. The primary aim of this chapter is to investigate the possible interaction between prolyl-tRNA synthetase (ProRS), the $tRNA^{Pro}$ charging domain of EPRS, and the HTLV-1 genome.

Materials and Methods

RNA Preparation

Seven different RNA constructs were used in these studies: human $tRNA^{Pro}$, HTLV-1 5'UTR, HTLV-1 μTLE, HTLV-1 TLE, HTLV-1 R, HTLV-1 U5-5'gag, and HTLV-1 5'gag. The unmodified human $tRNA^{Pro}$ UGG construct was used. The HTLV-1 5'UTR RNA consists of the first 633 nucleotides (nt) of the HTLV-1 viral RNA with a 15-nt 5' linker added. The HTLV-1 μTLE consists of the 19-nt TLE stem loop with the C-G base pair at the bottom of the stem mutated to G-C to improve transcription with T7 RNA polymerase. The HTLV-1 TLE is a 147-nt RNA that consists of the TLE stem loop and the surrounding structure. The HTLV-1 R, U5-5'gag, and 5'gag RNAs are fragments of the HTLV-1 5'UTR, each with the 15-nt 5' linker as well. R corresponds to the 5' terminal repeat sequence. U5-5'gag consists of the sequence between the beginning of the unique 5' region and the beginning of the gag coding region. This includes the entire TLE region and the PBS. Finally, the 5'gag sequence consists of the first 175-nt of the gag coding region.

To prepare the HTLV-1 μTLE RNA, single-stranded template oligomers with the T7 polymerase promoter and the RNA coding sequence were ordered from IDT. The sense and anti-sense oligomers were then annealed by heating to 80°C, cooling to 60°C, and cooling on ice. The μTLE RNA was *in vitro* transcribed from this double stranded template. The HTLV-1 5'UTR construct is contained in a pUC19 plasmid, and the transcription template can be obtained by restriction digest with BstYI (W. Cantara & W. Wu, unpublished data). The TLE, R, U5-5'gag, and 5'gag RNAs were cloned out of the full 5'UTR plasmid into pUC19 vectors behind T7 promoters (U5-5'gag, and 5'gag cloning performed by W. Cantara & W. Wu). The TLE vector was then digested with PstI, and the other three were digested with BstYI to obtain their transcription templates.

After *in vitro* transcription with T7 RNA polymerase, RNAs were gel purified via denaturing PAGE (8M urea). Excised bands were then crushed and incubated in RNA elution buffer (0.5 mM NH_4OAc, 1 mM EDTA) at 37°C overnight. Solutions were then butanol extracted and ethanol precipitated yielding pure RNA. RNAs were then quantified using UV-vis spectrometry. Extinction coefficients at 260 nm for HTLV-

1 RNAs are as follows: μTLE, 1.320 X 10^5 M^{-1} cm^{-1}; TLE, 1.328 X 10^6 M^{-1} cm^{-1}; R, 2.085 X 10^6 M^{-1} cm^{-1}; U5-5'gag, 2.160 X 10^6 M^{-1} cm^{-1}; 5'gag, 1.730 X 10^6 M^{-1} cm^{-1}.

Internally ^{32}P-labeled RNAs were *in vitro* transcribed with the same templates as unlabeled constructs in the presence of [α^{32}P] ATP. They were then purified as described for unlabeled RNAs and quantified using scintillation counting.

Protein Preparation

Human ProRS is the $tRNA^{Pro}$ charging domain of the multifunctional fusion protein EPRS, and is the primary construct used in this study. This construct was cloned from the EPRS sequence into a pKS509 vector, and transformed into BL21(DE3) cells as described in Heacock et al. (1996). Cultures were grown in media with 100 μg/μL ampicillin, and induced with 1 mM IPTG to overexpress ProRS. Recovered cells were lysed by sonication, and nucleic acids were precipitated with 5 mg/mL protamine sulfate. ProRS was then purified by Ni^{2+} affinity chromatography as described in Heacock et al. (1996). Protein elution fractions were then buffer exchanged into 2X storage buffer (50 mM HEPES pH 7.4, 300 mM NaCl, 2 mM β-mercaptoethanol) and diluted in 80% glycerol to achieve a final 40% glycerol storage solution. Concentration was determined by Bradford assay. LysRS-ΔN65 was purified as stated in Chapter 2 under "Protein purification."

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) were employed to investigate ProRS binding to the HTLV-1 5'UTR. Internally ^{32}P-labeled HTLV-1 5'UTR RNA was folded in 50 mM HEPES (pH 7.4) buffer by heating to 80°C for 2 min, cooling to 60°C for 2 min, adding 10 mM $MgCl_2$, incubating at 37°C for 30 min, and cooling on ice for 30 min. The folded RNAs were then incubated with serially diluted protein (concentrations ranging from 0 to 10 μM) at room temperature for 30 min. The proteins assayed were ProRS and LysRS-ΔN65. The reactions were run on native (1 mM $MgCl_2$) PAGE at 4°C and analyzed by phosphorimaging.

Fluorescent RNA labeling

Vicinal hydroxyl groups exclusive to the 3' terminus of RNAs (2.5 pmols) were oxidized with 1.5 mM sodium periodate in 100 mM sodium acetate (pH 5.1) solution resulting in two 3' terminal aldehydes (Jones et al., 2013). These reactions were ethanol precipitated after 1.5 hours. The resuspended RNAs were then 3' end-labeled with fluorescein-5-thiosemicarbozide (FTSC) by incubation with ten-fold excess FTSC in 100 mM sodium acetate buffer at 4°C overnight (Jones et al., 2013). There free dye was then removed with G-25 Roche spin columns, and the RNAs were assessed for purity using urea PAGE. Labeling efficiencies were determined by UV-vis spectrometry (FTSC ε_{495}=85000$M^{-1}cm^{-1}$).

Fluorescent Protein Labeling

Human ProRS was labeled with fluorescein isothiocyanate (FITC) by incubating 100 µM ProRS with 1 mM FITC at 4°C overnight. The free dye was removed by buffer exchanging with 2X ProRS storage buffer using 10 kDa MW cutoff centrifugal filters (Amicon). Solution was assessed for purity by SDS-PAGE. The labeling efficiencies were determined by UV-vis spectrometry (FITC ε_{495}=73000$M^{-1}cm^{-1}$)

Fluorescence anisotropy direct binding assay

For fluorescence anisotropy (FA) binding , FTSC-labeled RNAs were folded in 50 mM HEPES pH 7.4 by heating to 80°C for 2 min, cooling to 60°C for 2 min, adding 1 mM $MgCl_2$, incubating at 37°C for 30 min (if the RNA is longer than 150 nt), and cooling on ice for 30 min. RNAs (10 nM) were then incubated with serially diluted ProRS at room temperature for 30 min in FA binding buffer (20 mM Tris pH 8, 15 mM NaCl, 35 mM KCl, 1 mM $MgCl_2$) prior to FA measurement. Data were plotted and curves were fit to the Hill equation.

Results and Discussion

Unlike HIV-1, the packaging of the HTLV-1 RT primer has not been well characterized. In fact, the sum of the data concerning the HTLV-1 $tRNA^{Pro}$ primer can be found in just two studies. The first study reported a putative PBS to be the reverse complement of the 3' end of $tRNA^{Pro}$ (Seiki et. al., 1982). The

second study showed that HTLV-1 infection stimulates the production of a 3' fragment of $tRNA^{Pro}$, and that this fragment is capable of priming RT and is packaged into virions (Ruggero et. al., 2013). Furthermore, this study showed that the full-length $tRNA^{Pro}$ is packaged into virions as well. (Ruggero et. al., 2013). Therefore, to date the only information known about HTLV-1 RT priming is that at least the 3' end of $tRNA^{Pro}$ is required, and that multiple potential primers are packaged into virions. Nothing is known about how these primer RNAs are packaged or localized.

Recently, our lab proposed a preliminary secondary structure of the HTLV-1 5'UTR solved with multiple SHAPE reagents (Figure 17A, W. Wu, unpublished data). From this structure, a hairpin was identified that has high sequence homology to the $tRNA^{Pro}$ anticodon stem loop—this hairpin will be

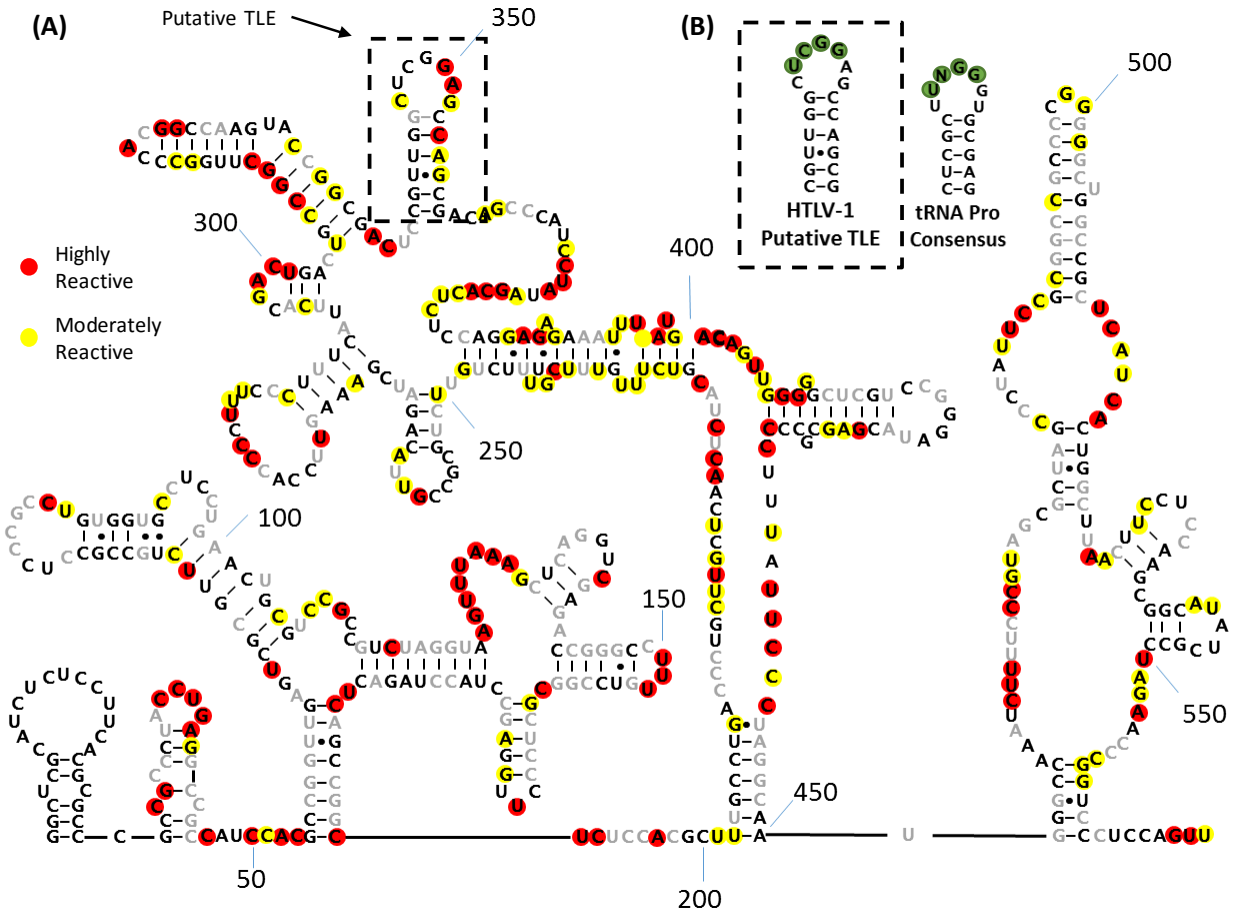


Figure 17: (A) The HTLV-1 5'UTR predicted secondary structure based on SHAPE. Red circles indicate high reactivity with 1M6 (>0.8), yellow circles indicate moderate reactivity (>0.4), and residues in grey text indicate areas with no reactivity data. (B) The hairpin in the black dashed box is the putative TLE, and the circles in green indicate the residues that are shared between it and the $tRNA^{Pro}$ anticodon stem loop consensus sequence.

Putative TLE Conservation

Compensatory Mutations

Mutation	Quantity	Percent
U342C	46	14.4%
G341A + U342C + C357U	7	2.2%
Total	**53**	**16.6%**

Interfering Mutations

Mutation	Quantity	Percent
G341A + U342C	6	1.9%
G341A + U342C + G356A	1	0.3%
G341C	1	0.3%
U347A	2	0.6%
G349A	2	0.6%
345-346insA + 356-357insA	1	0.3%
dG349	2	0.6%
dG349 + dC353	1	0.3%
dC354	1	0.3%
Total	**17**	**5.3%**

Innocuous Mutations

Mutation	Quantity	Percent
U342C + A355G	1	0.3%
C346G	1	0.3%
C348U	2	0.6%
C357U	1	0.3%
Total	**5**	**1.6%**

Summary

Sequence by Type	Qunatity	Percent
No Mutation	245	76.6%
Compensatory	53	16.6%
Interfering	17	5.3%
Innocuous	5	1.6%
Total	**320**	

Conservation by Nucleotide

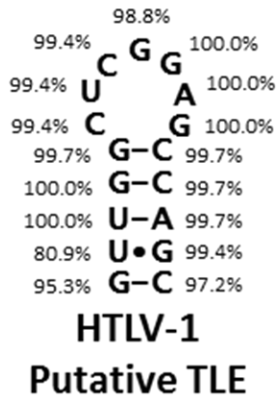


Figure 18: The putative HTLV-1 TLE is shown to be highly conserved when compared to 320 HTLV-1 LTR sequences from different groups and subgroups. These tables list every mutation found in the 17-nt stem and loop of the putative TLE. The vast majority of sequences are identical to the one used in *in vitro* experiments, and only 5.3% of sequences contain mutations that would negatively Impact the functionality of a TLE. Furthermore, only three individual nucleotides in the hairpin are less than 98% conserved: G341, U342, and C357. However, the majority of mutations at these nucleotides are compensatory mutations.

referred to as the putative TLE. The putative TLE contains one of the $tRNA^{Pro}$ anticodon sequences CGG at the center of a 7-nt loop that also contains a uridine residue in the correct position to form a canonical tRNA U-turn structure (Figure 17B). Additionally, this putative TLE is located just 5' to the PBS, which would make it optimal for primer localization. It also resembles the arrangement of functional elements in the HIV-1 5'UTR. Sequence analysis shows that the putative TLE is highly-conserved with three sequence variants accounting for 93% of all sequences, and only about 5% of sequences contain any kind of mutation that could inhibit this hairpin's function as a TLE (Figure 18). Multiple complex retroviruses package the aminoacyl-tRNA synthetase that binds to their tRNA primer, presumably to facilitate packaging of the primer (Cen et. al., 2002). As mentioned earlier, it has also been show that HIV-1 uses a TLE to facilitate primer localization (Jones et. al., 2013; Jones & Cantara et. al., 2014). Although it is not known whether HTLV-1 packages EPRS to facilitate $tRNA^{Pro}$ primer packaging, the existence of this putative TLE led to the hypothesis that HTLV-1 employs tRNA mimicry analogous to HIV-1 to localize $tRNA^{Pro}$ to the

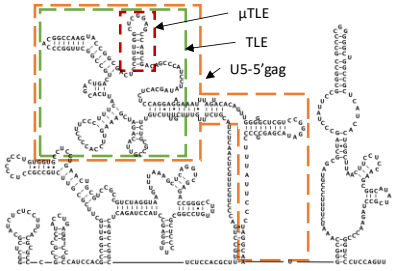
PBS. To test this hypothesis, various binding experiments were performed to test the interaction between ProRS and various RNA constructs of the HIV-1 genome.

Electrophoretic mobility shift assays (EMSAs) were used to assay ProRS binding to the full length HTLV-1 5'UTR. To control for specificity to ProRS, LysRS-ΔN65 was assayed in parallel. However, K_d values could not be calculated due to consistent RNA aggregation in the wells at higher concentrations of both proteins. Therefore, these assays were inconclusive. It is possible that this assay failed because the protein:RNA complex was too large to fit into the matrix of the gel. Alternatively, these RNA binding proteins could be causing this large RNA to aggregate.

The next binding scheme that was attempted was to titrate fluorescently-labeled protein with various RNA constructs including $tRNA^{Pro}$, and HTLV-1 R, U5-5'gag, and 5'gag. However, even at very high concentrations of RNA no change in anisotropy was observed. There are a few reasons that could explain the failure of this binding scheme. The first is that large globular proteins typically exhibit fairly high anisotropy in solution without ligand bound. This could make the change in anisotropy upon ligand binding imperceptible. This has not been ruled out, but it should be noted that the baseline anisotropy values that fluorescently labeled ProRS exhibited were not dissimilar to those exhibited by free tRNA in solution. Another potential explanation of these results lies in the nature of ProRS. As a class II tRNA synthetase, ProRS is functional as a dimer. Therefore, the low assay concentrations may be too far below the dimerization constant, and it may not efficiently bind to RNA. The final possible explanation comes from the process used to fluorescently label the protein. The isothiocyanate component of FITC is intended to react with free amine groups of proteins. However, amines from amino acids such as lysine can be important for interaction with nucleic acids in many proteins, so it is possible that the chemical labeling of ProRS is interfering with RNA binding.

The final binding scheme, and the only successful one, involved titrating fluorescently-labeled RNA constructs with ProRS. Although very high concentrations of the protein were needed to get

RNA	ProRS K_d (μM)	n
$tRNA^{Pro}$	2.30 ± 0.20	2.6 ± 0.1
μTLE	5.29 ± 0.36	2.7 ± 0.3
TLE	4.32 ± 0.29	2.0 ± 0.2
U5-5'gag	4.59 ± 0.62	1.6 ± 0.3



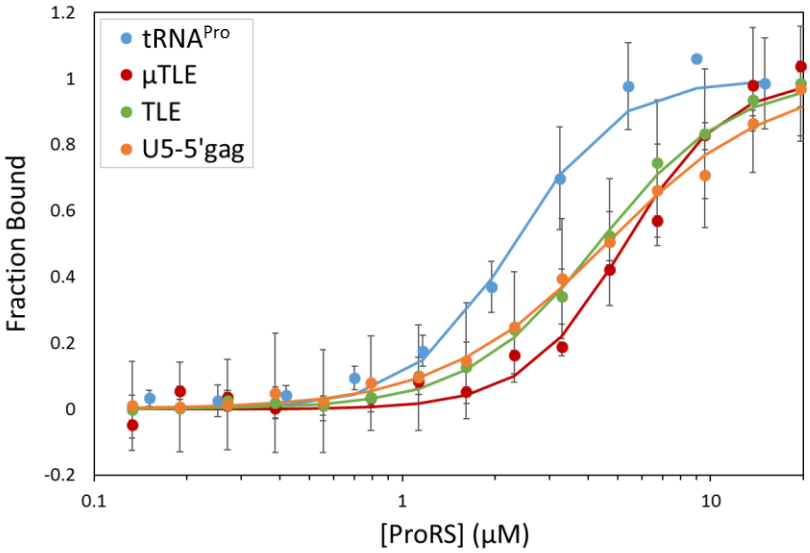


Figure 19: FA assays with titrated ProRS show that ProRS binds approximately two-fold tighter to $tRNA^{Pro}$ than to putative TLE containing RNAs. This suggests that the putative TLE would not bind to ProRS efficiently enough to compete with $tRNA^{Pro}$, rendering this hairpin ineffective for primer localization.

saturated binding curves (including for the tRNA), K_d values were able to be calculated for multiple constructs using this method: $tRNA^{Pro}$, μTLE, TLE, and U5-5'gag. The Hill equation was used to fit these plots, and calculate K_d and n values. It was found that ProRS binds consistently to RNAs containing the putative TLE with two-fold higher K_d values than to $tRNA^{Pro}$ [p-values <0.01] (Figure 19). This suggests that ProRS is not binding specifically to the putative TLE. This low specificity implies that the putative TLE would be unable to compete with the tRNA for binding to ProRS, making this hairpin a poor primer-localizing device. A negative control still needs to be completed to confirm this result, however preliminary experiments suggest the binding will be similar to that of RNAs containing the TLE.

There are a few obvious limitations to these experiments. First, it is possible that ProRS binds to the putative TLE more specifically in the context of other regions of the 5'UTR that are not included in any of these constructs. Second, ProRS is naturally a part of the bifunctional fusion protein, EPRS, and other regions of this protein that are not included in this ProRS construct could affect this domain's ability to bind RNA, and its specificity towards RNAs. Finally, it is possible that this hairpin is not functioning as a TLE in a manner similar to that of the HIV-1 TLE.

Conclusions and Future Directions

These results suggest that the putative TLE hairpin is not mimicking $tRNA^{Pro}$, and do not support the original hypothesis that the function of this hairpin is to localize the RT primer. This leaves us with the open questions: (1) In the absence of a TLE how does HTLV-1 localize its RT primer? and (2) What function does this highly-conserved hairpin serve? There are a few alternative hypotheses that still need to be explored to fully eliminate the possibility of this hairpin's involvement in tRNA localization. First, the ProRS domain may not reflect of the binding specificity of EPRS. Conversely, EPRS may not be the only protein that binds specifically to $tRNA^{Pro}$. It is also possible that the conservation of this hairpin is due to its involvement in a completely different function such as binding to Gag during genomic RNA packaging. Therefore, to further understand the mechanism of primer localization and other functions of HTLV-1 RNAs, additional studies need to be performed to survey the proteins that bind specifically to $tRNA^{Pro}$, and to the HTLV-1 genomic RNA.

Finally, to better understand the mechanism of RT primer packaging, *in vivo* experiments need to be performed to determine whether EPRS is involved in the life cycle of HTLV-1. To date, it remains unknown whether EPRS is packaged into HTLV-1 virions. Therefore, even if EPRS is not involved in primer localization, it still may be involved in primer packaging, similar to the proposed function of synthetases in HIV-1 and RSV (Cen et al., 2002). Alternatively, EPRS is a key factor in the formation of the interferon-gamma-activated inhibitor of translation (GAIT) complex, which functions to silence the translation of transcripts containing GAIT elements (Sampath et al., 2004). Furthermore, HIV-1 infection results in the down regulation of EPRS (Mohammadi et al., 2013) and it is hypothesized that this functions to minimize GAIT complex silencing of HIV-1 translation. It is not clear whether HTLV-1 has this same effect on EPRS. However, if this down regulation is shared between these viruses, this could explain why HTLV-1 would employ an alternative primer-localization mechanism that does not involve EPRS. Future experiments will

contribute to a better understanding of these lifecycle mechanisms that are crucial to a retrovirus such as HTLV-1.

Acknowledgements

I would like to thank Weixin Wu for performing the SHAPE experiments to solve the secondary structure of the HTLV-1 genome to make this entire project possible.

Conclusion

HIV-1 and HTLV-1 are deadly viruses and the only two retroviruses known to cause human disease. HIV-1, the known cause of AIDS, is a global pandemic infecting 37 million people (World Health Organization, 2015), while HTLV-1 infects from 5-20 million people and causes ATL, HAM/TSP, and other associated diseases in a fraction of the infected population. This emphasizes the importance of studying these viruses with the hope of gaining knowledge that can eventually be translated into treating these patients. In this thesis, studies in both viruses focused on the functionally-dense and highly-conserved 5'UTR of the viral genomic RNA were performed.

To understand the structure of a large RNA, the first step is to experimentally determine the secondary structure using various RNA probing methods. These methods can also be useful for probing the functionalities of these RNAs. RNA probing with RNases and SHAPE has been optimized for experimentation with several RNA constructs. Additionally, the analysis method that has been developed has built upon existing software, particularly QuShape (Karabiber et al., 2013) and FAST (Pang et al., 2011), to create a more accurate and easier-to-use analysis tool. This approach has allowed us to solve the 5'UTR secondary structure of HTLV-1.

One function of the 5'UTR that is of particular interest is that of tRNA localization. Previous studies have demonstrated that knockdown of LysRS in HIV-1 infected cells results in lower viral infectivity (Guo et al, 2003) because the synthetase is packaged into virions to facilitate tRNA packaging (Cen et al., 2001). LysRS is not only involved in this function of primer packaging, but it has also been established to be involved in primer localization to the PBS through interaction with the TLE (Jones et al., 2013; Jones & Cantara et al., 2014). Direct binding data reveals that LysRS binds tighter to the TLE in constructs containing two of the Psi stem loops than to the TLE alone (Jones et al., 2013). RNA probing is being employed to probe this interaction. Qualitatively analyzed RNase probing data suggest that LysRS is interacts with SL1 and SL2 of Psi, as well as with the TLE.

We are also interested in the overall structure of the 5'UTR of HIV-1. Previously, SAXS has been used to generate models of three individual pieces of the 5'UTR (Jones & Cantara et al., 2014). Additionally, a large structure of the minimal packaging signal has also been solved using NMR (Keane et al., 2015). Using SAXS, an envelope of the HIV-1 5'UTR(240) has been generated that includes all of the Keane et al. NMR structure plus the PBS/TLE region. This structure has shown that 3D L-shape of a tRNA found in Jones & Cantara et al. is conserved when the TLE is in the context of this larger RNA. Additionally, it was established that the inter-helical orientations of the NMR model in Keane et al. are likely incorrect. Finally, when this SAXS envelope is compared to the envelope of a similar region of the MAL isolate's 5'UTR, it can be seen that the 3D structure is conserved.

The process of primer localization has been explored in HIV-1. However, not much is known about this process, or that of primer packaging, in HTLV-1. It has been shown that HIV-1 and RSV package the aminoacyl-tRNA synthetase that matches their tRNA primers into virions (Cen et al., 2002). Although, HTLV-1 is a complex retrovirus like these other two retroviruses, it is not known whether EPRS is involved in tRNA packaging. The recently solved secondary structure of the HTLV-1 5'UTR revealed what appeared to be a TLE, and was situated in a similar location relative to the PBS as the TLE is in HIV-1. Additionally, sequence analysis determined that the putative TLE is highly-conserved. This led to the hypothesis that EPRS may be involved in the localization of the HTLV-1 primer. This hypothesis was tested *in vitro* using direct binding assays to the ProRS domain of EPRS. It was demonstrated that the putative TLE-containing RNAs bound with two-fold higher K_d values to ProRS, suggesting that EPRS may not be involved in the localization of the tRNA to the HTLV-1 PBS. Future studies will address the potential role of other domains of EPRS in this binding interaction.

References

Akbarin, M. M., Rahimi, H., HassanNia, T., Razavi, G. S., Sabet, F., Shirdel, A. (2013). Comparison of HTLV-1 Proviral Load in Adult T Cell Leukemia/Lyphoma (ATL), HTLV-1-Associated Mylopathy (HAM-TSP), and Healthy Carriers. Iranian Journal of Basic Medical Sciences, 16:208-212.

Arts, E. J., & Hazuda, D. J. (2012). HIV-1 Antiretroviral Drug Therapy. Cold Spring Harbor Perspectives in Medicine, 2(4).Drucker, E., Alcabes, P. G., & Marx, P. A. (2001). The injection century: Massive unsterile injections and the emergence of human pathogens. The Lancet, 358(9297), 1989-1992.

Aviran, S., Trapnell, C., Lucks, J. B., Mortimer, S. A., Luo, S., Schroth, G. P., Doudna, J., Arkin A., Pachter, L. (2011). Modeling and automation of sequencing-based characterization of RNA structure. Proceedings of the National Academy of Sciences, 108(27), 11069-11074.

Bannwarth, S., & Gatignol, A. (2005). HIV-1 TAR RNA: The Target of Molecular Interactions Between the Virus and its Host. Current HIV Research, 3(1), 61-71.

Basu, V. P., Song, M., Gao, L., Rigby, S. T., Hanson, M. N., & Bambara, R. A. (2008). Strand transfer events during HIV-1 reverse transcription. Virus Research, 134(1-2), 19-38.

Biggar, R., Ng, J., Kim, N., Hisada, M., Li, H., Cranston, B., Hanchard, B., Maloney, E. (2006). Human Leukocyte Antigen Concordance and the Transmission Risk via Breast-Feeding of Human T Cell Lymphotropic Virus Type I. The Journal of Infectious Diseases J INFECT DIS, 193(2), 277-282.

Cen, S., Javanbakht, H., Kim, S., Shiba, K., Craven, R., Rein, A., Ewalt, K., Schimmel, P., Musier-Forsyth, K., & Kleiman, L. (2002). Retrovirus-Specific Packaging of Aminoacyl-tRNA Synthetases with Cognate Primer tRNAs. Journal of Virology, 76(24), 13111-13115.

Cook, L. B., Elemans, M., Rowan, A. G., & Asquith, B. (2013). HTLV-1: Persistence and pathogenesis. Virology, 435(1), 131-140.

De Thé, G., & Bomford, R. (1993). An HTLV-I Vaccine: Why, How, for Whom? AIDS Research and Human Retroviruses, 9(5), 381-386.

Freed, E. O., & Martin, A. M. (2007). Human Immunodeficiency Viruses: Replication. In Fields, B. N., Knipe, D. M., & Howley, P. M. (2007). Fields virology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Fujii, M., & Matsuoko, M. (2007). Human T-Cell Leukemia Virus Type 1 and 2. In Fields, B. N., Knipe, D. M., & Howley, P. M. (2007). Fields virology. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Gallo, R. C., & Montagnier, L. (2003). The Discovery of HIV as the Cause of AIDS. New England Journal of Medicine, 349(24), 2283-2285.

Gallo, R. C. (2005). History of the discoveries of the first human retroviruses: HTLV-1 and HTLV-2. Oncogene, 24(39), 5926-5930.

Gessain, A., & Cassar, O. (2012). Epidemiological Aspects and World Distribution of HTLV-1 Infection. Frontiers in Microbiology, 3, 388

Goldschmidt V., Paillart J. C., Rigourd M., Ehresmann B., Aubertin A. M., Ehresmann C., Marquet R. (2004). Structural variability of the initiation complex of HIV-1 reverse transcription. Journal Biological Chemistry 279: 35923–35931.

Goncalves, D. U., Proietti, F. A., Ribas, J. G., Araujo, M. G., Pinheiro, S. R., Guedes, A. C., & Carneiro-Proietti, A. B. (2010). Epidemiology, Treatment, and Prevention of Human T-Cell Leukemia Virus Type 1-Associated Diseases. Clinical Microbiology Reviews, 23(3), 577-589.

Gao F., Robertson D. L., Carruthers C. D., Li Y., Bailes E., Kostrikis L. G., Salminen M. O., Bibollet-Ruche F., Peeters M., Ho D. D., et al. (1998). An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I). Journal of Virology 72: 10234–10241.

Guo, F., Cen, S., Niu, M., Javanbakht, H., & Kleiman, L. (2003). Specific Inhibition of the Synthesis of Human Lysyl-tRNA Synthetase Results in Decreases in tRNALys Incorporation, tRNA3LysAnnealing to Viral RNA, and Viral Infectivity in Human Immunodeficiency Virus Type 1. Journal of Virology, 77(18), 9817-9822.

Guo, F., Gabor, J., Cen, S., Hu, K., Mouland, A. J., & Kleiman, L. (2005). Inhibition of Cellular HIV-1 Protease Activity by Lysyl-tRNA Synthetase. Journal of Biological Chemistry, 280(28), 26018-26023.

Heacock, D., Forsyth, C. J., Shiba, K., & Musier-Forsyth, K. (1996). Synthesis and Aminoacyl-tRNA Synthetase Inhibitory Activity of Prolyl Adenylate Analogs. Bioorganic Chemistry, 24(3), 273-289.

Helga-Maria C., Hammarskjöld M. L., Rekosh D. (1999). An intact TAR element and cytoplasmic localization are necessary for efficient packaging of human immunodeficiency virus type 1 genomic RNA. Journal of Virology 73(5):4127–4135.

Hiroyuki, S., & Susumu, K. (1981). New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). Gene, 16(1-3), 73-78.

Irish, B. P., Khan, Z. K., Jain, P., Nonnemacher, M. R., Pirrone, V., Rahman, S., Rajagopalan, N., Suchitra, J. B., Mostoller, K., Wigdahl, B. (2009). Molecular Mechanisms of Neurodegenerative Diseases Induced by Human Retroviruses: A Review. American Journal of Infectious Diseases, 5(3), 231-258.

Jiang, M., J. Mak, A. Ladha, E. Cohen, M. Klein, B. Rovinski, and L. Kleiman. (1993). Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. J. Virol. 67:3246–3253.

Jones, C. P., Saadatmand, J., Kleiman, L., & Musier-Forsyth, K. (2013). Molecular mimicry of human tRNALys anti-codon domain by HIV-1 RNA genome facilitates tRNA primer annealing. RNA, 19(2), 219-229.

Jones, C. P., Cantara, W. A., Olson, E. D., & Musier-Forsyth, K. (2014). Small-angle X-ray scattering-derived structure of the HIV-1 5′ UTR reveals 3D tRNA mimicry. Proceedings of the National Academy of Sciences, 111(9), 3395-3400.

Karabiber, F., Mcginnis, J. L., Favorov, O. V., & Weeks, K. M. (2012). QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. Rna, 19(1), 63-73.

Keane, S. C., Heng, X., Lu, K., Kharytonchyk, S., Ramakrishnan, V., Carter, G., Barton, S., Hosic, A., Florwick, A., Santos, J., Bolden, N. C., McCowin S., Case D. A., Johnson, B. A., Salemi, M., Telensnitsky, A., Summers, M. F. (2015). Structure of the HIV-1 RNA packaging signal. Science, 348(6237), 917-921.

Kchour, G., Tarhini, M., Kooshyar, M., Hajj, H. E., Wattel, E., Mahmoudi, M., Hatoum, H., Rahimi, H., Maleki, M., Rafatpanah, H., Rezaee, S. A. R., Yazdi, M. T., Shirdel, A., deThe, H., Hermine, O., Farid, R., & Bazarbachi, A. (2009). Phase 2 study of the efficacy and safety of the combination of arsenic trioxide, interferon alpha, and zidovudine in newly diagnosed chronic adult T-cell leukemia/lymphoma (ATL). Blood, 113(26), 6528-6532.

Konarev P. V., Petoukhov M. V., Volkov V. V., Svergun D. I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. Journal Appl Cryst 39:277–286.

Kovaleski, B. J., Kennedy, R., Hong, M. K., Datta, S. A., Kleiman, L., Rein, A., & Musier-Forsyth, K. (2006). In Vitro Characterization of the Interaction between HIV-1 Gag and Human Lysyl-tRNA Synthetase. Journal of Biological Chemistry, 281(28), 19449-19456.

Lairmore, M. D., Haines, R., & Anupam, R. (2012). Mechanisms of human T-lymphotropic virus type 1 transmission and disease. Current Opinion in Virology, 2(4), 474-481.

Lu K., Heng X., Summers M. F. (2011). Structural determinants and mechanism of HIV-1 genome packaging. J Mol Biol 410(4):609–633.

Mak, J., & Kleiman, L. (1997). Primer tRNAs for Reverse Transcription. Journal of Virology, 71(11), 8087-8095

Marquet, R., Isel, C., Ehresmann, C., & Ehresmann, B. (1995). TRNAs as primer of reverse transcriptases. Biochimie, 77(1-2), 113-124.

Matutes, E., Taylor, G. P., Cavenagh, J., Pagliuca, A., Bareford, D., Domingo, A., Hamblin, M., Kelsey, S., Mir, N., Reilly, J. T. (2001). Interferon alpha and zidovudine therapy in adult T-cell leukaemia lymphoma: Response and outcome in 15 patients. British Journal of Haematology Br J Haematol, 113(3), 779-784.

Meiering, C. D., & Linial, M. L. (2001). Historical Perspective of Foamy Virus Epidemiology and Infection. Clinical Microbiology Reviews, 14(1), 165-176.

Merino, E. J., Wilkinson, K. A., Coughlan, J. L., & Weeks, K. M. (2005). RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). Journal of the American Chemical Society, 127(12), 4223-4231.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Research, 15(21), 8783-8798.

Mohammadi, P., Desfarges, S., Bartha, I., Joos, B., Zangger, N., Muñoz, M., Günthard, H., Beerenwinkel, N., Telenti, A., Ciuffi, A. (2013). 24 Hours in the Life of HIV-1 in a T Cell Line. PLoS Pathogens, 9(1).

O'Haver, T. (2016). A Pragmatic Introduction to Signal Processing. University of Maryland at College Park.

Osame, M. (2002). Pathological mechanisms of human T-cell lymphotropic virus type I-associated myelopathy (HAM/TSP). Journal of Neurovirology, 8(5), 359-364.

Pang, P. S., Elazar, M., Pham, E. A., & Glenn, J. S. (2011). Simplified RNA secondary structure mapping by automation of SHAPE data analysis. Nucleic Acids Research, 39(22).

Peckham, C., & Gibb, D. (1995). Mother-to-Child Transmission of the Human Immunodeficiency Virus. New England Journal of Medicine, 333(5), 298-303.

Rice, G. M., Leonard, C. W., & Weeks, K. M. (2014). RNA secondary structure modeling at consistent high accuracy using differential SHAPE. Rna, 20(6), 846-854.

Ruggero, K., Guffanti, A., Corradin, A., Sharma, V. K., Bellis, G. D., Corti, G., Grassi, A., Zanovello, P., Bronte, V., Ciminale, V., & D'agostino, D. M. (2014). Small Noncoding RNAs in Cells Transformed by Human T-Cell Leukemia Virus Type 1: A Role for a tRNA Fragment as a Primer for Reverse Transcriptase. Journal of Virology, 88(7), 3612-3622.

Seiki, M., Hattori, S., & Yoshida, M. (1982). Human adult T-cell leukemia virus: Molecular cloning of the provirus DNA and the unique terminal structure. Proceedings of the National Academy of Sciences, 79(22), 6899-6902.

Shiba, K., Stello, T., Motegi, H., Noda, T., Musier-Forsyth, K., & Schimmel, P. (1997). Human Lysyl-tRNA Synthetase Accepts Nucleotide 73 Variants and Rescues Escherichia coli Double-defective Mutant. Journal of Biological Chemistry, 272(36), 22809-22816.

Skripkin E., Paillart J. C., Marquet R., Ehresmann B., Ehresmann C. (1994). Identification of the primary site of the human immunodeficiency virus type 1 RNA dimerization in vitro. Proceedings National Academy of Science USA 91(11):4945–4949.

Vasa, S. M., Guex, N., Wilkinson, K. A., Weeks, K. M., & Giddings, M. C. (2008). ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. Rna, 14(10), 1979-1990.

Weeks, K. M. (2010). Advances in RNA structure analysis by chemical probing. Current Opinion in Structural Biology, 20(3), 295-304.

World Health Organization. (2015). HIV/AIDS Fact Sheet (Rep. No. 360).

Xu, Z., & Culver, G. M. (2009). Chemical Probing of RNA and RNA/Protein Complexes. Methods in Enzymology Biophysical, Chemical, and Functional Probes of RNA Structure, Interactions and Folding: Part A, 147-165.